

Title

Collectin-complement activating protein chimeras

Field of invention

5

The present invention relates to a fusion protein capable of activating the complement system, methods for producing said fusion protein as well as pharmaceutical composition comprising said fusion protein and methods for treating diseases, in particular infections, with said fusion protein.

10

Background of invention

15

Animals have developed different complex strategies to protect themselves against infections. The immune responses can be divided into two main groups, the adaptive immune response, in which an adaptation has taken place and in which cells play a dominant part and the innate immune response, which is available instantly and which primarily is based on molecules present in the body fluids. The innate immune system is operational at time of birth, in contrast to the adaptive immune defence which only during infancy obtains its full power of protecting the body (Janeway *et al.*, 1999).

20

25

Bacteria entering the body at mucosal surfaces or through broken skin are immediately recognised by collectins, a family of soluble proteins that recognise distinctive carbohydrate configurations that are present on the surfaces of microbes and absent from the cells of the multicellular organism. Collectins thus belong to the large and diverse group of pattern recognition receptors of the innate immune system. In humans, three collectins are known, although others may exist: cows for example have more. Collectins target the particles to which they bind either for uptake by phagocytes or for activation of the complement cascade, and in these ways can mediate their destruction.

30

35

Collectins all exhibit the following architecture: they have an N-terminal cysteine-rich region that appears to form inter-chain disulfide bonds, followed by a collagen-like region, an α -helical coiled-coil region and finally a C-type lectin domain which is the pattern-recognizing region and is referred to as the carbohydrate recognition domain

(CRD). The name collectin is derived from the presence of both collagen and lectin domains. The α -helical coiled-coil region initiates trimerisation of the individual polypeptides to form collagen triple coils, thereby generating collectin subunits each consisting of 3 individual polypeptides, whereas the N-terminal region mediates formation of oligomers of subunits. Different collectins exhibit distinctive higher order structures, typically either tetramers of subunits or hexamers of subunits. The grouping of large numbers of binding domains allows collectins to bind with high avidity to microbial cell walls, despite a relatively low intrinsic affinity of each individual CRD for carbohydrates.

C-type CRDs are found in proteins with a widespread occurrence, both in phylogenetic and functional perspective. The different CRDs of the different collectins enable them to recognise a range of distinct microbial surface components exposed on different microorganisms. The terminal CRDs are distributed in such a way that all three domain target surfaces that present binding sites has a spacing of approximately 53 Å (Sheriff *et al.*, 1994; Weis & Drickamer, 1994). This property of 'pattern recognition' may contribute further to the selectively binding of microbial surfaces. The collagenous region or possibly the N-terminal tails of the collectins, are recognised by specific receptors on phagocytes, and is the binding site for associated proteases that are activated to initiate the complement cascade upon binding of the CRD domain to a target.

Mannan-binding lectin (MBL) also termed mannose-binding lectin or mannose binding protein is a collectin which has gained great interest as an important part of the innate immune system. MBL binds to specific carbohydrate structures found on the surface of a range of microorganisms including bacteria, yeast, parasitic protozoa and viruses, and has been found to exhibit antibacterial activity through killing mediated by activation of the terminal, lytic complement components or through promotion of phagocytosis. MBL deficiency is associated with susceptibility to frequent infections by a variety of microorganisms in childhood, and possibly also in adults.

The CRD of MBL recognises preferentially hexoses with equatorial 3- and 4-OH groups, such as mannose, glucose, *N*-acetylmannosamin and *N*-acetyl glucoseamin while carbohydrates which do not fulfil this sterical requirement, such as galactose and D-fucose, are not bound (Weis *et al.*, 1992). The carbohydrate selectivity is ob-

viously an important aspect of the self/non-self discrimination by MBL and is probably mediated by the difference in prevalence of mannose and *N*-acetyl glucoseamin residues on microbial surfaces, one example being the high content of mannose in the cell wall of yeasts such as *Saccharomyces cerevisiae* and *Candida albicans*.

5 Carbohydrate structures in glycosylation of mammalian proteins are usually completed with sialic acid, which prevents binding of MBL to these oligomeric carbohydrates and thus prevents MBL recognition of 'self' surfaces. Also, the trimeric structure of each MBL subunit may be of importance for target recognition.

10 Complement is a group of proteins present in blood plasma and tissue fluid that aids the body's defences following an infection. The complement system is being activated through at least three distinct pathways, designated the classical pathway, the alternative pathway, and the MBlectin pathway (Janeway *et al.*, 1999). The classical pathway is initiated when complement factor 1 (C1) recognises surface-bound
15 immunoglobulin. The C1 complex is composed of two proteolytic enzymes, C1r and C1s, and a non-enzymatic part, C1q, which contains immunoglobulin-recognising domains. C1q and MBL shares structural features, both molecules having a bouquet-like appearance when visualised by electron microscopy. Also, like C1q, MBL is found in complex with two proteolytic enzymes, the mannan-binding lectin associated proteases (MASP). The three pathways all generate complement factor 3 (C3)
20 convertase, which ensures the binding of C3b to the surface of the activating surface, *i.e.* the targeted microbial pathogen. Conversion of C3 into surface bound C3b is pivotal in the process of eliminating the microbial pathogen by phagocytosis or lysis (Janeway *et al.*, 1999).

25 Certain O-antigen specific oligosaccharides of *Salmonella* have been reported to activate complement in C4-deficient guinea-pig serum and *Salmonella* serogroup C was later shown to react with MBL and hence activate complement by the MBlectin pathway, which is also termed the MBL pathway of complement activation or the
30 lectin pathway.

It has for some time been speculated that the innate immune system may collaborate with the adaptive immune system in the generation of specific immune responses as exemplified by the antibody response after infection or vaccination.

35 Fearon's group have shown that the attachment of the C3d fragment of complement

factor C3 onto a protein antigen through fusion by gene technology can increase the immunogenicity of the antigen 1000 fold or more. Practical applications of this technique, or any number of modifications, are still awaited.

5 The importance of the complement system for normal immune responses was first suggested by Pepys, who found impaired antibody responses to sheep erythrocytes, a thymus-dependent antigen, in mice that were C3-depleted with cobra venom factor. The idea of a link between innate and adaptive immunity was supported by reports demonstrating reduced primary antibody responses to thymus dependent anti-
10 gens and impaired IgM to IgG switching in patients and experimental animals with deficiencies of C4, C2 and C3. The mechanism may involve the generation of C3-derived ligands for binding of antigen or antigen-containing complexes to complement receptors on B lymphocytes or antigen-presenting cells. Thus, blocking of CR1 (CD35) and CR2 (CD21) in mice with specific anti CR1 and anti-CR2 antibodies or
15 with soluble receptor protein reduced antibody responses to immunisation and experiments with CR1 and CR2-deficient knock-out mice show the requirement of these receptors for responses to thymus-dependent antigens. In addition, patients with leucocyte adhesion deficiency, who lack the CD11/CD18 adhesion molecule CR3, demonstrate impaired antibody responses and failure to switch from IgM to
20 IgG. The C3-derived fragment C3d, a specific CR2 ligand, as mentioned above, show a strong dose-dependent adjuvant effect.

Deficiencies of the classical complement pathway (C1, C2, C4 and C3) are associated with infections by encapsulated bacteria. The main reason for this is probably
25 the reduced efficiency of opsonic and bactericidal defence mechanisms caused by complement dysfunction. However, impaired immune responses to polysaccharide antigens might also be considered. The influence of complement on responses to thymus-independent antigens has not been extensively studied, and the available information is contradictory. Thus, low antibody responses to thymus-independent
30 antigens have been clearly documented in C3-depleted mice and C3-deficient dogs. On the other hand, some reports find that C3-deficient patients appear to respond normally to immunisation with polysaccharide vaccines.

Ficolins, like MBL, are lectins that contain a collagen-like domain. Unlike MBL, however, they have a fibrinogen-like domain, which is similar to fibrinogen β - and γ -
35

chains. Ficolins also forms oligomers of structural subunits, each of which is composed of three identical 35 kDa polypeptides. Each subunit is composed of an amino-terminal, cysteine-rich region; a collagen-like domain that consists of tandem repeats of Gly-Xaa-Yaa triplet sequences (where Xaa and Yaa represent any amino acid); a neck region; and a fibrinogen-like domain. The oligomers of ficolins comprises two or more subunits, especially a tetrameric form of ficolin has been observed.

Some of the ficolins triggers the activation of the complement system substantially in similar way as done by MBL. This triggering of the complement system results in the activation of novel serine proteases (MASPs) as described above.

The fibrinogen-like domain of several lectins has a similar function to the CRD of C-type lectins including MBL, and hereby function as pattern-recognition receptors to discriminate pathogens from self.

Serum ficolins have a common binding specificity for GlcNAc (N-acetylglucosamine), elastin or GalNAc (N-acetyl-galactosamine). The fibrinogen-like domain is responsible for the carbohydrate binding. In human serum, two types of ficolin, known as L-ficolin (P35, ficolin L, ficolin 2 or hucolin) and H-ficolin (Hakata antigen, ficolin 3 or thermolabile b2-macroglycoprotein), have been identified, and both of them have lectin activity. L-ficolin recognises GlcNAc and H-ficolin recognises GalNAc. Another ficolin known as M-ficolin (P35-related protein, Ficolin 1 or Ficolin A) is not considered to be a serum protein and is found in leucocytes and in the lungs. L-ficolin and H-ficolin activate the lectin-complement pathway in association with MASPs. M-Ficolin, L-ficolin and H-ficolin has calcium-independent lectin activity.

MASPs (MBL-associated serine proteases) comprising MASP-1, MASP-2 and MASP-3 are proteolytic enzymes that are responsible for activation of the lectin pathway. The overall structure of MASPs resembles that of the two proteolytic components of the first factor in the classical complement pathway, C1r and C1s. The lectin pathway is initiated when MBL or a ficolin associated with MASP-1, MASP-2, MASP-3 and sMAP binds to a carbohydrate structure of the surfaces of e.g. bacteria, yeast, parasitic protozoa, viruses. MASP-2 is the enzyme component that – like

C1s in the classical pathway – cleaves the complement components C4 and C2 to form the C3 convertase C4bC2a, which is common to both the lectin- and classical-pathway activation routes.

5 MASP-1, MASP-2, MASP-3 and sMAP are encoded by two genes; sMAP is a truncated form of MASP-2, and MASP-3 is produced from the MASP-1 gene by alternative splicing. The MASP-1 gene has an H-chain-encoding region that is common to MASP-1 and MASP-3, which is followed by tandem repeats of protease-domain-encoding regions that are specific to MASP-3 and MASP-1.

10

The MASP family can be divided into two phylogenetic lineages – TCN-type and AGY-type lineages. The TCN-type lineage, which includes MASP-1, has a TCN codon (where N denotes A, G, C or T) that encodes the active-site serine, the presence of a histidine-loop disulphide bridge and split exons. By contrast, the AGY-type
15 lineage, which includes MASP-2, MASP-3, C1r and C1s, is characterised by an AGY codon (where Y denotes C or T) that encodes the active-site serine, the absence of a histidine-loop and a single exon.

15

20 MASP-1, MASP-2, MASP-3, C1r and C1s consist of six domains: two C1r/C1s/Uegf/bone morphogenetic protein 1 (CUB) domains; an epidermal growth factor (EGF)-like domain; two complement control protein (CCP) domains or short consensus repeats (SCRs), and a serine-protease domain. Histidine (H), aspartic acid (D) and serine (S) residues are essential for the formation of the active centre in the serine-protease domain. Only MASP-1 has two additional cysteine residues in
25 a light chain, which form a histidine-loop disulphide bridge (S-S), as is found in trypsin and chymotrypsin. On binding of MBL and ficolins to carbohydrate on the surface of a pathogen, the pro-enzyme form of a MASP is cleaved between the second CCP and the protease domain, which results in an active form that consists of two polypeptides – heavy and light chains (also known as A and B chains).

25

30

Summary of invention

The present invention relates to fusion proteins capable of activating the complement system. Accordingly, the present invention relates to a fusion protein comprising
35 ing

35

a first polypeptide sequence derived from a lectin-complement pathway activating protein (=complement activating protein) or a functional homologue thereof; and
a second polypeptide sequence derived from a collectin or a functional homologue thereof;

wherein said complement activating protein is not a collectin.

The fusion protein is suitable for use in treatment consisting of creation, reconstitution, enhancing and/or stimulating the opsonic and/or bactericidal activity of the complement system, i.e. enhancing the ability of the immune defence to recognise and kill microbial pathogens, and accordingly, the invention relates to a medicament comprising the fusion protein.

Also, in another aspect the invention relates to a method of treatment of a clinical condition in an individual in need thereof comprising administering to said individual the fusion protein as defined above.

In another aspect the invention relates to a method of treatment or prophylaxis of a clinical condition, such as infection, in an individual in need thereof comprising administering to said individual a the fusion protein a first polypeptide sequence derived from a protein capable of forming oligomers of structural units; and a second polypeptide sequence derived from a mannose binding lectin (MBL, wherein said first polypeptide sequence and said second peptide sequence is not derived from the same protein, and said fusion protein is capable of associating with mannose-associated serine protease (MASP). The first polypeptide sequence is preferably derived from a protein capable of forming tetramers, pentamers, and/or hexamers of a structural unit. In a preferred embodiment the first polypeptide sequence and the second polypeptide sequence are as described below.

In a further aspect the invention relates to use of the fusion protein as defined above for the preparation of a medicament for the treatment of a clinical condition in an individual in need thereof.

Furthermore the invention relates to a method for producing the fusion protein, as well as an isolated nucleic acid sequence encoding the fusion protein, a vector comprising the sequence and a cell comprising the vector.

5 Drawings

Figure 1 shows the sequence of L ficolin

Figure 2 shows the sequence of MBL

Figure 3 shows an example of a fusion protein

10 Figures 4-8 show plasmids as described in Example 1

Figure 9 shows an alignment of fusion proteins described in Example 1

Figure 10 shows two Western blots as discussed in Example 2.

Definitions

15

Collectins: A family of structurally related, carbohydrate-recognising proteins of innate immunity, including mannan-binding lectin (MBL) and surfactant proteins A and D. The name refers to the presence of a collagen-like region and a C-type lectin domain.

20

Complement: A group of proteins present in blood plasma and tissue fluid that aids the body's defences following an infection. Complement is involved in destroying foreign cells and attracting phagocytes to the area of conflict in the body.

25

Conjugated: An association formed between two compounds for example between an immunogenic determinant and a collectin and/or collectin homologue or between an immunogenic determinant and a saccharide. The association may be a physical association generated e.g. by the formation of a chemical bond, such as e.g. a co-valent bond.

30

CRD: Carbohydrate recognition domain, a C-type lectin domain that is found at the C-terminus of collectins.

35

Immunogenic determinant: A molecule, or a part thereof, containing one or more epitopes that will stimulate the immune system of a host organism to make a secre-

tory, humoral and/or cellular antigen-specific response, or a DNA molecule which is capable of producing such an immunogen in a vertebrate.

5 Immune response: Response to an immunogenic composition comprising an immunogenic determinant. An immune response involves the development in the host of a cellular- and/or humoral immune response to the administered composition or vaccine in question. An immune response generally involves the action of one or more of i) the antibodies raised, ii) B cells, iii) helper T cells, iv) suppressor T cells, v) cytotoxic T cells and iv) complement directed specifically or unspecifically to an
10 immunogenic determinant present in an administered immunogenic composition.

Lectin: Proteins that specifically bind carbohydrates.

15 MASP: Mannose-associated serine protease

MBL: Mannan-binding lectin or mannose-binding lectin.

Subunit complex=structural unit: complex of three individual fusion proteins, like the subunit complex discussed above for MBL and ficolins.

20

Detailed description of the invention

An object of the present invention is to provide a fusion protein capable of activating the complement system in order to aid in preventing or treating diseases, in particular infectious diseases.
25

The fusion protein is composed of

30 a first polypeptide sequence derived from a lectin-complement pathway activating protein (=complement activating protein) or from a functional homologue thereof; and

a second polypeptide sequence derived from a collectin or from a functional homologue thereof;
35

wherein said complement activating protein is not a collectin.

By combining a polypeptide sequence derived from a lectin-complement pathway activating protein and a polypeptide sequence derived from a collectin it is possible to design a fusion protein having binding affinity for a variety of carbohydrates, preferably bacterial and viral carbohydrates and at the same time having complement system activating activity.

First polypeptide sequence

10

The first polypeptide sequence may be derived from any lectin-complement pathway activating protein. Said lectin-complement pathway activating protein may be naturally occurring lectin-complement pathway activating protein as well as variants or homologues to said lectin-complement pathway activating proteins, wherein said variants or homologues have maintained the lectin-complement pathway activating activity.

15

It is preferred that the fusion protein is capable of forming subunit complexes, each consisting of 3 individual fusion proteins as defined above.

20

Also the first polypeptide sequence is preferably capable of forming oligomeric complexes with the first polypeptide sequence of another fusion protein, wherein said another fusion protein may be identical to the first fusion protein. Thereby an oligomeric complex of two or more fusion proteins or two or more subunit complexes may be provided, said oligomeric complex having a higher binding avidity for bacterial or viral carbohydrates than the monomeric fusion protein. In a preferred embodiment the oligomeric complex is a dimeric subunit complex, more preferably a trimeric subunit complex, more preferably a tetrameric subunit complex.

25

In a preferred embodiment the lectin-complement pathway activating protein is a ficolin as defined above. Said ficolin may be L-ficolin, H-ficolin or M-ficolin or variants or homologues thereof. In a preferred embodiment the lectin-complement pathway activating protein is L-ficolin.

30

In another embodiment the first polypeptide sequence comprises the fibrinogen domain of the lectin, and/or the neck region of a lectin, such as a ficolin or a homologue or a variant thereof.

5 When the first polypeptide sequence is derived from ficolin or from a variant or a homologue of ficolin, it is preferred that the first polypeptide sequence comprises the collagen-like domain from ficolin or from a variant or homologue of ficolin. In another embodiment it is preferred that the first polypeptide sequence comprises the Cystein rich domain from ficolin or from a variant or homologue of ficolin. It is even more
10 preferred that the first polypeptide sequence comprises the collagen-like domain and the Cystein rich domain from ficolin or from a variant or homologue of ficolin.

It is more preferred that the first polypeptide sequence comprises the collagen-like domain from L-ficolin or from a variant or homologue of L-ficolin. In another em-
15 bodiment it is more preferred that the first polypeptide sequence comprises the Cystein rich domain from L-ficolin or from a variant or homologue of L-ficolin. It is even more preferred that the first polypeptide sequence comprises the N-terminal region of L-ficolin including two Cystein amino acid residues.

20 It is even more preferred that the first polypeptide sequence comprises the collagen-like domain and the Cystein rich domain from L-ficolin or from a variant or homologue of L-ficolin.

In a particular preferred embodiment the ficolin has one of the sequences listed be-
25 low with reference to their database and accession No. For each of the sequences the Cystein rich region and the collagen-like region is described.

30 NP_003656. ficolin 3 precursor; ficolin (collagen/fibrinogen domain-containing) 3 (Hakata antigen) [Homo sapiens] [gi:4504331]

90..299 /region_name="pfam00147, fibrinogen_C, Fibrinogen beta and gamma chains, C-terminal globular domain"
35 90..299 /region_name="smart00186, FBG, Fibrinogen-related domains (FReDs); Domain present at the C-termini of fibrinogen beta and gamma chains, and a variety of fibrinogen-related proteins, including tenascin and Drosophila scabrous"

1 mdllwilpsl wlllggpac lktqehpscp gpreleaskv vllpscpgap gspgekgapg

61 pqqpppppgk mgpkgepgdp vnllrcqegp rncrellsqg atlsqwyhlc lpegralpvf
 121 cdmdtegggw lvfqrqdgds vdffrswssy ragfgnqese fwlgnenlhq ltlqgnwelr
 181 veledfngnr tfahyatfrl lgevdhyqla lgkfsegtag dslslhsgrp ftydadhd
 241 snsncavivh gawwyascyr snlmgryavs daaahkygid wasgrgvghp yrrvmmlr

5

XP_116792. similar to Ficolin 2 precursor (Collagen/fibrinogen domain-containing protein 2) (Ficolin-B) (Ficolin B) (Serum lectin P35) (EBP-37) (Hucolin) (L-Ficolin) [Homo sapiens] [gi:20477458]

10

91..168 /region_name="pfam00147, fibrinogen_C, Fibrinogen beta and gamma chains, C-terminal globular domain"
 91..168 /region_name="smart00186, FBG, Fibrinogen-related domains (FReDs); Domain present at the C-termini of fibrinogen beta and gamma chains, and a variety of fibrinogen-related proteins, including tenascin and Drosophila scabrous"

15

1 mgpallalsf lwtmaltedt cpamleyval nsepgmaskn psrrhglsl vvdqpgparg
 61 vrtddqpsga dpgslelhge cpifpsequi lthhnnypfs tedqdndrda encavhyqga
 121 wwyaschls ingvylggar dsftnginwk sgkgnnysyk vsemkvrpt

20

O00602. Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) [gi:20455484]

25

1..29 /gene="FCN1" /region_name="Signal" /note="POTENTIAL."
 30..326 /gene="FCN1" /region_name="Mature chain" /note="FICOLIN 1."
 55..93 /gene="FCN1" /region_name="Domain" /note="COLLAGEN-LIKE."
 133 /gene="FCN1" /region_name="Conflict" /note="T -> N (IN REF. 1)."
 144..290 /gene="FCN1" /region_name="Domain" /note="FIBRINOGEN C-TERMINAL."
 287 /gene="FCN1" /region_name="Conflict" /note="N -> S (IN REF. 1)."
 305 /gene="FCN1" /site_type="glycosylation" /note="N-LINKED (GLCNAC...) (POTENTIAL)."

30

35

1 melsgatmar glavlvlfl hiknlpaqaa dtcpevkvg legsdktlil rgcpglpgap
 61 gpkgeagvig ergerglpga pgkagpvgpk gdrgekmgmrg ekgdagqsqs catgprnckd
 121 lldrgyflsg whtiylpdcrl ptylclmdt dgggwtvfr rmdgsvdfyr dwaaykqgfg
 181 sqlgefwlgn dnihaltagg sselrvdlvd fegnhqfaky ksfkvadeae kyklvlgafv
 241 ggsagnsltg hnnnffstkd qdndvssnc aekfqgawwy adchasing lylmgphesy
 301 anginwsaak gykysykvse mkvrpa //

40

O75636. Ficolin 3 precursor (Collagen/fibrinogen domain-containing protein 3) (Collagen/fibrinogen domain-containing lectin 3 P35) (Hakata antigen) [gi:13124185]

45

1..21 /gene="FCN3" /region_name="Signal" /note="POTENTIAL."
 22..299 /gene="FCN3" /region_name="Mature chain" /note="FICOLIN 3."
 48..80 /gene="FCN3" /region_name="Domain" /note="COLLAGEN-LIKE."
 50 /gene="FCN3" /site_type="hydroxylation"
 53 /gene="FCN3" /site_type="hydroxylation"
 59 /gene="FCN3" /site_type="hydroxylation"
 65 /gene="FCN3" /site_type="hydroxylation"
 68 /gene="FCN3" /site_type="hydroxylation"
 77 /gene="FCN3" /site_type="hydroxylation"

50

119..265 /gene="FCN3" /region_name="Domain" /note="FIBRINOGEN C-
TERMINAL."

189 /gene="FCN3" /site_type="glycosylation" /note="N-LINKED (GLCNAC...) (PO-
TENTIAL)."

1 mdllwilpsl wllllggpac lktqehpscp gpreleaskv vllpscpgap gspgekgapg
61 pqgppgppgk mgpkgepgdp vnllrcqegp mcrellsqg atlsqwyhlc lpegalpvf
121 cdmtdteggw lvfqrqdg vdffrswssy ragfgnqese fwlgnehlq ltlqgnwelr
181 veledfngnr tfahyatfrl lgevdhyqla lgkfsegtag dslslhsgp ftydadhdh
241 snscavivh gawwyascyr snlngryavs daaahkygid wasgrgvghp yrrvrmmlr

XP_130120. similar to Ficolin 2 precursor (Collagen/fibrinogen domain-containing
protein 2) (Ficolin-B) (Ficolin B) (Serum lectin P35) (EBP-37) (Hucolin) [Mus mus-
culus] [gi:20823464]

59..95 /region_name="Collagen triple helix repeat (20 copies)" /note="Collagen"
/db_xref="CDD:pfam01391"

59..89 /region_name="Collagen triple helix repeat (20 copies)" /note="Collagen"
/db_xref="CDD:pfam01391"

60..95 /region_name="Collagen triple helix repeat (20 copies)" /note="Collagen"
/db_xref="CDD:pfam01391"

60..95 /region_name="Collagen triple helix repeat (20 copies)" /note="Collagen"
/db_xref="CDD:pfam01391"

60..95 /region_name="Collagen triple helix repeat (20 copies)" /note="Collagen"
/db_xref="CDD:pfam01391"

60..95 /region_name="Collagen triple helix repeat (20 copies)" /note="Collagen"
/db_xref="CDD:pfam01391"

60..95 /region_name="Collagen triple helix repeat (20 copies)" /note="Collagen"
/db_xref="CDD:pfam01391"

61..95 /region_name="Collagen triple helix repeat (20 copies)" /note="Collagen"
/db_xref="CDD:pfam01391"

61..95 /region_name="Collagen triple helix repeat (20 copies)" /note="Collagen"
/db_xref="CDD:pfam01391"

61..95 /region_name="Collagen triple helix repeat (20 copies)" /note="Collagen"
/db_xref="CDD:pfam01391"

103..312 /region_name="Fibrinogen beta and gamma chains, C-terminal globular
domain" /note="fibrinogen_C" /db_xref="CDD:pfam00147"

103..312 /region_name="Fibrinogen-related domains (FReDs)" /note="FBG"
/db_xref="CDD:smart00186"

1 malgsaalfv lltvhaagt cpeklvidle gykqltilqg cpplpgaagp kgeagakgdr
61 gesglpgipg kegptgpkgn qgekgrgek gdsgpsqscs tgprtkell tqghfltgwy
121 tiylpdcrlp tvlcmdtdg ggwtvfqrrl dgsvdffrdw tsykrfgsq lgefvlgnnd
181 ihalttqgts elrvdlsdfe gkhdfakyss fqiqqeaeay klilgnflgg gagdsltphn
241 nrlfstkdqd ndgstsscam gyhgawwysq chtsnlmgly lrgphksyan gvnwkswrgy
301 nysckvsemk vrlr

NP_056654. ficolin 2 isoform d precursor; ficolin (collagen/fibrinogen domain-
containing lectin) 2 (hucolin); ficolin (collagen/fibrinogen domain-containing lectin) 2;
hucolin [Homo sapiens] [gi:8051590]

39..95 /region_name="collagen-like domain"

- 1 meldravglv gaatlillsfl gmawalqaad tcpevkmgvl egskltilr gcpglpgapg
61 dkgeagtnkg rgergppgpp gkagppgpng apgeppclt gd
- 5 NP_056653. ficolin 2 isoform c precursor; ficolin (collagen/fibrinogen domain-containing lectin) 2 (hucolin); ficolin (collagen/fibrinogen domain-containing lectin) 2; hucolin [Homo sapiens] [gi:8051588]
- 39..95 /region_name="collagen-like domain"
- 10 102..143 /region_name="Fibrinogen beta and gamma chains, C-terminal globular domain" /note="fibrinogen_C" /db_xref="CDD:pfam00147"
- 102..143 /region_name="Fibrinogen-related domains (FReDs)" /note="FBG" /db_xref="CDD:smart00186"
- 15 1 meldravglv gaatlillsfl gmawalqaad tcpevkmgvl egskltilr gcpglpgapg
61 dkgeagtnkg rgergppgpp gkagppgpng apgeppclt gprtckdlld rghflsgwht
121 iylpdcrlt vlcdmtdgg gwtvsvglr ggqpgspggq aahlvgehtl efsillvgds
181 qr
- 20 NP_056652. ficolin 2 isoform b precursor; ficolin (collagen/fibrinogen domain-containing lectin) 2 (hucolin); ficolin (collagen/fibrinogen domain-containing lectin) 2; hucolin [Homo sapiens] [gi:8051586]
- sig_peptide 1..25
- 25 mat_peptide 26..275
- 60..275 /region_name="FBG domain" /note="fibrinogen beta/gamma homology"
- 64..275 /region_name="Fibrinogen-related domains (FReDs)" /note="FBG" /db_xref="CDD:smart00186"
- 64..274 /region_name="Fibrinogen beta and gamma chains, C-terminal globular domain" /note="fibrinogen_C" /db_xref="CDD:pfam00147"
- 30 1 meldravglv gaatlillsfl gmawalqaad tcpgergppg ppgkagppgp ngapgeppc
61 ltgprtckdl ldrghflsgw htiylpdcrlp ltvcdmtd gggwtvfqrr vdgsvdfyrd
121 watykqgfgs rlgefvlgn d nihaltaggt selrvdlvdf ednyqfakyr sfkvadeaek
35 181 ynlvlgafve gsagdsllfh nnqsfstkdg dndlntgnca vmfqqawwyk nchvsnlng
241 ylrghthsfa nginwkskg ynysykvsem kvrrpa
- NP_001994. ficolin 1 precursor; ficolin (collagen/fibrinogen domain-containing) 1 [Homo sapiens] [gi:8051584]
- 40 sig_peptide 1..27
- mat_peptide 28..326
- 40..108 /region_name="collagen-like domain"
- 50..105 /region_name="Collagen triple helix repeat (20 copies)" /note="Collagen" /db_xref="CDD:pfam01391"
- 45 51..107 /region_name="Collagen triple helix repeat (20 copies)" /note="Collagen" /db_xref="CDD:pfam01391"
- 52..106 /region_name="Collagen triple helix repeat (20 copies)" /note="Collagen" /db_xref="CDD:pfam01391"
- 50 115..326 /region_name="FBG domain" /note="fibrinogen beta/gamma homology"
- 115..326 /region_name="Fibrinogen-related domains (FReDs)" /note="FBG" /db_xref="CDD:smart00186"

15

115..325 /region_name="Fibrinogen beta and gamma chains, C-terminal globular domain" /note="fibrinogen_C" /db_xref="CDD:pfam00147" variation 315
 /db_xref="dbSNP:1128428" variation 316 /db_xref="dbSNP:1128429" variation 317
 /db_xref="dbSNP:1128430"

5

1 melsgatmar glavllvfl hiknlpaqaa dtcpevkvg legsdkltl rgcpplpgap
 61 gpkgeagvig ergerlpga pgkagpvgpk gdrgekmg rg ekgdagqsqs catgprnckd
 121 lldrgyflsg whtiylpdcrl pltlcdmdt dgggwtvfqr rmdgsvdfyr dwaaykqgfg
 181 sqlgefwn dnhaltagq sselrvldv fegnhqfaky ksfkvadeae kyklvlgafv
 241 ggsagnsltg hnnffstkd qndvssnc aekfgawwy adchasnlg lymgphesy
 301 anginwsaak gykyskvse mkvrpa

10

NP_004099. ficolin 2 isoform a precursor; ficolin (collagen/fibrinogen domain-containing lectin) 2 (hucolin); ficolin (collagen/fibrinogen domain-containing lectin) 2; hucolin [Homo sapiens] [gi:4758348]

15

sig_peptide 1..25

mat_peptide 26..313

39..95 /region_name="collagen-like domain"

20

98..313 /region_name="FBG domain" /note="fibrinogen beta/gamma homology"

102..313 /region_name="Fibrinogen-related domains (FReDs)" /note="FBG"

/db_xref="CDD:smart00186"

102..312 /region_name="Fibrinogen beta and gamma chains, C-terminal globular domain" /note="fibrinogen_C" /db_xref="CDD:pfam00147"

25

1 melldragvvl gaatlslfl gmawalqaad tpevkmgvl egskltlir gcpplpgapg
 61 dkgeagtngk rgerpppgpp gkagpppgng apgeppclt gprtkdld rhflsgwh
 121 iylpdcrlpl vlcdmdtdgg gwtvfqrrvd gsvdfyrdwa tykqgfsrl gefwlgndni
 181 haltaggtse lrvldvdfed nyqfakysf kvadeaekyn lvgafvegs agdsltfnh
 241 qsfstkddn dlntgncavm fggawwyknc hvsnlngryl rgthgsfang inwkskgyn
 301 ysykvsemkv rpa

30

Q9WTS8. Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) [gi:13124116]

35

1..22 /gene="FCN1" /region_name="Signal" /note="POTENTIAL."

23..335 /gene="FCN1" /region_name="Mature chain" /note="FICOLIN 1."

50..88 /gene="FCN1" /region_name="Domain" /note="COLLAGEN-LIKE."

152..298 /gene="FCN1" /region_name="Domain" /note="FIBRINOGEN C-

40

TERMINAL."

271 /gene="FCN1" /site_type="glycosylation" /note="N-LINKED (GLCNAC...) (POTENTIAL)."

45

1 mwwpmlwafp vlclcssqa lqesgacpd vkivglgaqd kvaviscps fpghpgpkge
 61 pgspagrger glqgspgkmg ppgskgepgt mgppgvkgek gergtasplg qkelgdalcr
 121 rgrpsckdll trgfltgwy tiylpdcrlpl tvcdmdvdg ggwtvfqrrv dgsinfyrdw
 181 dsykrfgnl gtefwlgndy lhltangnq elrvdlrefq gqtsfakys fqvsgaqeky
 241 klitqgfleg tagdsltahn nmafsthdd ndtnggknc alfhwawwyh dchqsnlgr
 301 ylpghesya dginwlsgrg hrysikvaem kiras

50

Q15485. Ficolin 2 precursor (Collagen/fibrinogen domain-containing protein 2) (Ficolin-B) (Ficolin B) (Serum lectin P35) (EBP-37) (Hucolin) (L-Ficolin) [gi:13124203]

1..25 /gene="FCN2" /region_name="Signal" /note="POTENTIAL."
 26..313 /gene="FCN2" /region_name="Mature chain" /note="FICOLIN 2."
 54..92 /gene="FCN2" /region_name="Domain" /note="COLLAGEN-LIKE."
 131..277 /gene="FCN2" /region_name="Domain" /note="FIBRINOGEN C-
 5 TERMINAL."
 240 /gene="FCN2" /site_type="glycosylation" /note="N-LINKED (GLCNAC...) (PO-
 TENTIAL)."
 300 /gene="FCN2" /site_type="glycosylation" /note="N-LINKED (GLCNAC...) (PO-
 10 TENTIAL)."

1 meldravglv gaatlslfl gmawalqaad tpevkmgvl egskdltlr gcpplpgapg
 61 dkgeagtnkg rgerpppgpp gkagppgpng apgeppclt gprtckdlld rghflsgwht
 121 iylpdcrlpt vlcdmtdgg gwtvfqrrvd gsvdfyrdwa tykqgfsrl gefwlgndni
 181 haltaagtse lrvdlvdfed nyqfakysf kvadeaekyn lvgafvegs agdsltfnh
 15 241 qsfstkqdn dlntgncavm fggawwyknc hvsnlngryl rgthgsfang inwksgkgyn
 301 ysykvsemky rpa

O70497. Ficolin 2 precursor (Collagen/fibrinogen domain-containing protein 2) (Fi-
 colin-B) (Ficolin B) (Serum lectin P35) (EBP-37) (Hucolin) [gi:13124181]

<1..15 /gene="FCN2" /region_name="Signal" /note="POTENTIAL."
 16..>306 /gene="FCN2" /region_name="Mature chain" /note="FICOLIN 2."
 41..79 /gene="FCN2" /region_name="Domain" /note="COLLAGEN-LIKE."
 130..276 /gene="FCN2" /region_name="Domain" /note="FIBRINOGEN C-
 25 TERMINAL."
 299 /gene="FCN2" /site_type="glycosylation" /note="N-LINKED (GLCNAC...) (PO-
 TENTIAL)."

1 lgsaalfvt ltvhaagtcp elkvidlegv kqltlqgcp glpgaagpkg eagakgdrge
 30 61 sglpgipgke gptgpkgnqg ekgirgekdg sgpsqscatg prtckelltg ghfltgwyti
 121 ylpdcrlptv lcdmtdggg wtvfqrldg svdfrrdws ykrfgsqglg efwlgndnih
 181 alttqgtsel rvdlsdfegk hdfakysf iqgeaekyl ilgnflggga gdsltphnrr
 241 lfstkdqnd gstsscamgy hgawwysqch tsnlnglyl rphksyangv nwkswrgyny
 301 sckvse

O70165. Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Fi-
 colin-A) (Ficolin A) (M-Ficolin) [gi:13124179]

1..22 /gene="FCN1" /region_name="Signal" /note="POTENTIAL."
 40 23..334 /gene="FCN1" /region_name="Mature chain" /note="FICOLIN 1."
 50..88 /gene="FCN1" /region_name="Domain" /note="COLLAGEN-LIKE."
 152..298 /gene="FCN1" /region_name="Domain" /note="FIBRINOGEN C-
 TERMINAL."
 261 /gene="FCN1" /site_type="glycosylation" /note="N-LINKED (GLCNAC...) (PO-
 45 TENTIAL)."

1 mqwptlwafs glclcpsqa lqergacpd vkvvlgaqd kvvviqscpg fpgppgpkge
 61 pgspagrger gfgsgpgkmg pagskgepgt mgppgvkgek gdtgaapslg ekelgdtlcq
 121 rgrpsckdl trglftgwy thlpdcrl tvcdmvdg ggwtvfqrrv dgsidfrdw
 181 dsykrfgnl gtefwlgndy lhlitangnq elrvldqdfq gkgsyakyss fqvseeqeky
 50 241 kltlqqlfeg tagdsltahn nmsftthdqd ndansmncaa lfhgawwyhn chqsnlrgy
 301 lsgshesyad ginwgtgqgh hysykvaemk iras

P57756. Ficolin 2 precursor (Collagen/fibrinogen domain-containing protein 2) (Ficolin-B) (Ficolin B) (Serum lectin P35) (EBP-37) (Hucolin) [gi:13124114]

1..22 /gene="FCN2" /region_name="Signal" /note="POTENTIAL."
 23..319 /gene="FCN2" /region_name="Mature chain" /note="FICOLIN 2."
 48..86 /gene="FCN2" /region_name="Domain" /note="COLLAGEN-LIKE."
 137..283 /gene="FCN2" /region_name="Domain" /note="FIBRINOGEN C-TERMINAL."
 306 /gene="FCN2" /site_type="glycosylation" /note="N-LINKED (GLCNAC...) (POTENTIAL)."

1 mvlgsaalfv lslcvteftl haadtcepvk vldlegsnkl tilqgcpglp galgpkgeag
 61 akgdrgesgl pghpgkagpt gpkgdrgckg vrgekgtgtp sqscatgprt ckelltrgyf
 121 ltgwytiylp dcrpltlvcd mtdgggwtv frridgtvd ffrdwtsykq gfgsqlgefz
 181 lgnndihalt tqgtnehrvd ldfdgndhdf akysffiqg eaekykllg nllgggagds
 241 ltsqnnmlfs tkdqndqgs sncavryhga wwysdchsn lnglylrglh ksyangvnwk
 301 swkgynysyk vsemkvrl

JC5980: ficolin-A precurs - mouse [gi:7513652]

1..21 /region_name="domain" /note="signal sequence"
 50..64 /region_name="domain" /note="collagen-like"
 68..106 /region_name="domain" /note="collagen-like"
 123..334 /region_name="domain" /note="fibrinogen beta/gamma homology #label FBG"

1 mqwptlwafs glclcpsqa lgqergacpd vkvvglaqd kvvviqscpg fpgppgpkge
 61 pgspagrger gfggspgkmg pagskgepgt mgppgvkgek gdtgaapslg ekelgdticq
 121 rgprscdkll trgfiltgwy tlhlpdcrpl tlcdmdvdg ggwtvfqrrv dgsidffrdw
 181 dsykrfgnl gtefwlgndy lhlitangnq elrvldqdfq gkgsyakyss fqvseeqeky
 241 klitgqfleg tagdsltikh nmsftthdqd ndansmncaa lfhgawwyhn chqsnlngry
 301 lsgshesyad ginwgtgqgh hysykvaemk iras

S61517. ficolin-1 precurs- human [gi:2135116]

1..326 /note="36K HLA-cross-reactive plasma protein; hucolin, 35K"
 1..22 /region_name="domain" /note="signal sequence"
 52..108 /region_name="region" /note="collagen-like"
 115..326 /region_name="domain" /note="fibrinogen beta/gamma homology #label FBG"
 305 /site_type="binding" /note="carbohydrate (Asn) (covalent)"

1 melsgatmar glavllvfl hiknlpaqaa dtcepvkvvg legsdktlil rgcpglpgap
 61 gpkgeagvig ergerglpga pgkagpvgpk gdrgekmgrg ekgdagqsqs catgprnckd
 121 lldrgyflsg whniylpdcrl pltlcdmdt dgggwtvfqr rmdgsvdfyr dwaaykqgfg
 181 sqlgefvlgn dnihaltagq sselrvdlvd fegnhqfaky ksfkvadeae kyklvlgafv
 241 ggsagnsltg hnnffstkd qdndvssnc aekfggawwy adchasslng lylmgphesy
 301 anginwsaak gykysykvse mkvrpa

A47172. transforming growth factor-beta 1-binding protein homolog ficolin-alpha - pig [gi:423206]

112..323 /region_name="domain" /note="fibrinogen beta/gamma homology #label FBG"

1 mdtgrvaaam rplvllvaf ctaapaldtc pevkvvgleg sdklsilrgc pglpgaagpk
 61 geagasgpgk gggppgapge pgppgpgkdr gekgepgpgk esweteqclt gprrckellt
 121 rghilsgwht iylpdcqplt vlcdmtdtgg gwtvfqrrsd gsvdfyrdwa aykrfgsq
 181 gefwlgndhi haltaqgtne lrvdlvdfeg nhqfakyrfs qvadeaekym lvgafvegn
 5 241 agdsltshnn slfttkdqn dqyasncavl yggawwynsc hvsningryl ggshgsfang
 301 vnwssgkgyn ysykvsemkf rat

JC4942. ficolin-1 precursor – human [gi:2135117]

10 1..22 /region_name="domain" /note="signal sequence"
 45..101 /region_name="region" /note="collagen-like"
 108..319 /region_name="domain" /note="fibrinogen beta/gamma homology #label
 FBG"
 111..315 /region_name="region" /note="fibrinogen-like".
 15 298 /site_type="binding" /note="carbohydrate (Asn) (covalent)"

1 marglavllv lflhiknlpa qaadtcevk vvglegsdki tilrgcpglp gapgpkgeag
 61 vigergergl pgapgkagpv gpkgdrgek mrgekdgagq sqscatgprn ckdlldrgyf
 121 lsgwhitiylp dcrplvlcd mtdtgggwtv fqrmdgsvd fyrdwaaykq gfgsqggef
 20 181 lgnndihalt aqgsselrvd lvdfeqnhqf akyksfkvad eaekykylvg afvggsagns
 241 ltghnnnffs tkdqndvss sncaekfqa wwyadchasn lnglylmgph esyanginws
 301 aakgykysk vsemkvrpa

AAF44911. symbol=BG:DS00929...[gi:7287873]

25 1 mkscffvfl wtlfevgs sphtcpsgsp ngihqlmlpe eepfqvtqck ttardwiviq
 61 rldgsvfn qswfsykdgfdp ngeffig lqklylmtr qphelfiqlk hpggatvyah
 121 fddfqvds etelyklervgk ysgtagdlsr yhinkrfstf drdndesskn caehgggww
 181 fhsclsr

30
 35 The first polypeptide preferably comprises at least 10, such as at least 12, for exam-
 ple at least 15, such as at least 20, for example at least 25, such as at least 30, for
 example at least 35, such as at least 40, for example at least 50 consecutive amino
 acid residues of the complement activating protein or of a variant or a homologue to
 said protein. Such a variant or homologue is preferably at least 70%, such as 80%,
 for example 90%, such as 95% identical to the complement activating protein.

40 The first polypeptide sequence of the fusion protein is preferably capable of activat-
 ing the lectin-complement pathway when bound directly or indirectly to a target,
 such as a bacteria or a virus. In a preferred embodiment the first polypeptide se-
 quence is capable of associating with at least one MASP protein, such as a MASP
 protein selected from the group consisting of MASP-1, MASP-2 and MASP-3 or
 functional homologues or variants hereof. In particular the first polypeptide is capa-
 45 ble of associating with said at least one MASP protein when being part of the fusion

protein. Thereby the first polypeptide sequence is capable of providing the fusion protein with complement system activating activity.

5 In a particular preferred embodiment the first polypeptide sequence comprises at least the amino acid residues corresponding to 1-54 of L-ficolin sequence of Figure 1, such as 1-55 of L-ficolin sequence of Figure 1, such as 1-69 of L-ficolin sequence of Figure 1, such as 1-77 of L-ficolin sequence of Figure 1, such as 1-90 of L-ficolin sequence of Figure 1, such as 1-93 of L-ficolin sequence of Figure 1, such as 1-131 of L-ficolin sequence of Figure 1, such as 1-207 of L-ficolin sequence of Figure 1. In
10 particular the first polypeptide sequence comprises the amino acid residues selected from: 1-55 of L-ficolin sequence of Figure 1, 1-54 of L-ficolin sequence of Figure 1, 1-50, or 1-77 of L-ficolin sequence of Figure 1. In a more preferred embodiment the first polypeptide sequence has the amino acid residues selected from: 1-55 of L-ficolin sequence of Figure 1, 1-54 of L-ficolin sequence of Figure 1, 1-50, or 1-77 of
15 L-ficolin sequence of Figure 1. In another embodiment the first polypeptide sequence comprises at least the amino acid residues corresponding to 60-90 of L-ficolin sequence of Figure 1, such as 55-90 of L-ficolin sequence of Figure 1, such as 54-92 of L-ficolin sequence of Figure 1.

20 It is preferred the first polypeptide sequence and the second polypeptide sequence are selected to include the motif X-X-G-X-X-G at least 5 times, such as at least 7 times, preferably in a consecutive sequence. It is more preferred to select the first polypeptide sequence and the second polypeptide sequence so that the aforementioned motif is substituted once with the motif X-X-G-X-G. In the motifs X means any
25 amino acid different from Glycine, and G means Glycine.

Second polypeptide sequence

30 The second polypeptide sequence is preferably capable of associating with one or more carbohydrates. This may be accomplished by incorporating at least the carbohydrate recognizing domain of the collectin in question. Accordingly, the second polypeptide sequence preferably comprises the CRD domain of a collectin or a homologue or a variant thereof.

Preferably the collectin is selected from the group consisting of MBL (mannose-binding lectin), SP-A (lung surfactant protein A), SP-D (lung surfactant protein D), BK (or BC, bovine conglutinin) and CL-43 (collectin-43). Most preferably the collectin is MBL.

In a particular preferred embodiment the collectin has one of the sequences listed below with reference to their database and accession No.

Collectins

SEQ ID NO: 42

Q9NPY3 Complement component C1q receptor precursor (Complement component 1, q subcomponent, receptor 1) (C1qRp) (C1qR(p)) (C1q/MBL/SPA receptor) (CD93 antigen) (CDw93) gi|21759074|sp|Q9NPY3|CD93_HUMAN[21759074]

FEATURES Location/Qualifiers source 1..652 /organism="Homo sapiens"

/db_xref="taxon:9606"

gene 1..652 /gene="C1QR1" /note="CD93"

Protein 1..652 /gene="C1QR1" /product="Complement component C1q receptor precursor"

Region 1..21 /gene="C1QR1" /region_name="Signal"

Region 22..652 /gene="C1QR1" /region_name="Mature chain"

/note="COMPLEMENT COMPONENT C1Q RECEPTOR."

Region 22 /gene="C1QR1" /region_name="Conflict" /note="T -> V (IN AA SEQUENCE)."

Region 24..580 /gene="C1QR1" /region_name="Domain" /note="EXTRACELLULAR (POTENTIAL)."

Region 32..174 /gene="C1QR1" /region_name="Domain" /note="C-TYPE LECTIN."

Region 36 /gene="C1QR1" /region_name="Conflict" /note="C -> T (IN AA SEQUENCE)."

Region 38..39 /gene="C1QR1" /region_name="Conflict" /note="TA -> RI (IN AA SEQUENCE)."

Region 155 /gene="C1QR1" /region_name="Conflict" /note="S -> N (IN REF. 1)."

Region 186 /gene="C1QR1" /region_name="Conflict" /note="G -> A (IN AA SEQUENCE)."

Region 260..301 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 1."

Bond bond(264,275) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILARITY."

Bond bond(271,285) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILARITY."

Bond bond(287,300) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILARITY."

Region 302..344 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 2."

Bond bond(306,317) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILARITY."

Bond bond(311,328) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILARITY."

Region 318 /gene="C1QR1" /region_name="Variant" /note="V -> A.
/FTId=VAR_013573."
Site 325 /gene="C1QR1" /site_type="glycosylation" /note="N-LINKED (GLCNAC...)
(POTENTIAL)."

5 Bond bond(330,343) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILAR-
ITY."

Region 345..384 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 3,
CALCIUM-BINDING (POTENTIAL)."

10 Bond bond(349,358) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILAR-
ITY."

Bond bond(354,367) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILAR-
ITY."

Bond bond(369,383) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILAR-
ITY."

15 Region 385..426 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 4,
CALCIUM-BINDING (POTENTIAL)."

Bond bond(389,400) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILAR-
ITY."

Bond bond(396,409) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILAR-
ITY."

20 Bond bond(411,425) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILAR-
ITY."

Region 427..468 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 5,
CALCIUM-BINDING (POTENTIAL)."

25 Bond bond(431,443) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILAR-
ITY."

Bond bond(439,452) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILAR-
ITY."

Bond bond(454,467) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILAR-
ITY."

30 Region 492 /gene="C1QR1" /region_name="Conflict" /note="S -> A (IN AA SE-
QUENCE)."

Region 496 /gene="C1QR1" /region_name="Conflict" /note="R -> Q (IN AA SE-
QUENCE)."

35 Region 504 /gene="C1QR1" /region_name="Conflict" /note="R -> G (IN AA SE-
QUENCE)."

Region 541 /gene="C1QR1" /region_name="Conflict" /note="P -> S (IN REF. 1)."

Region 581..601 /gene="C1QR1" /region_name="Transmembrane region"
/note="POTENTIAL."

40 Region 594..601 /gene="C1QR1" /region_name="Domain" /note="POLY-LEU."

Region 602..652 /gene="C1QR1" /region_name="Domain" /note="CYTOPLASMIC
(POTENTIAL)."

ORIGIN 1 matsmgllll lllltppga gtgadteavv cvgtacytah sgklsaaeeq nhcnqnggnl
61 atvkskeeeq hvqrvlaql rreaaltarm skfwiglqre kgkldpslp lkgfswvggg
45 121 edtpysnwhk elrnsciskr cvslldlsq plpsrlpkw segpcgspgs pgsniegfv
181 kfsfkgmcrp lalggpgqvt ytpfqtss sleavpfasa anvacgegdk detqshyflc
241 kekadvfdw gssgplcvsp kygcfnngg chqdcfeggd gsflcgrpg frliddlvtc
301 asrnpccssp crggatcvlg phgknytrc pqgyqldssq ldcvdvdecq dspcaqecvn
361 tpggfrcecw vgyepggpge gacqdvdeca lgrspcaqgc tntdgsfhcs ceegyvlage
50 421 dgtqcqdvde cvpgggplcd slcfntqgsf hcgcplgwwl apngvscmtg pvsllgppsgp
481 pdeedkgeke gstvpraata sprtpegt katptsrps lssdapitsa plkmilapsgs
541 pgvwrepsih hataasgpqe paggdssvat qnndgtdgqk lllyilgtv vaillllala
601 lgllyvrkr akreekkkk pqnaadsysw vperaesram enqysptpgt dc

SEQ ID NO: 43
 BAC05523 collectin placenta 1 [Mus musculus]
 gi|21901969|dbj|BAC05523.1|[21901969]
 5 FEATURES Location/Qualifiers source 1..742 /organism="Mus musculus"
 /db_xref="taxon:10090" /tissue_lib="Liver"
 Protein 1..742 /product="collectin placenta 1"
 CDS 1..742 /gene="CL-P1" /coded_by="AB078434.1:92..2320"
 ORIGIN 1 mkddfaeeee vqsfgykrfg iqegtqctkc knnwalkfsi vlyilcall titvailgyk
 10 61 vvekmdnvt d gmetshqtyd nklavesdl kklgdqagkk alstnselst frsdildlrq
 121 qlqeiteks knkdtleklq angdslvdrq sqlketlqnn sflittvnkt lqayngyvt n
 181 lqqdtsvlqg nlqsqmysqs vvimnlnln ltqvqqmli snlqqsvddt slaiqriknd
 241 fnlqqvflq akkdt dwlke kvqslqtlaa nnsalakann dtledmnsq l ssftgqmdni
 301 ttisqaneqs lkdldlhkd tenrtavkfs qleerfqvfe tdvniisni sytahhlrt l
 15 361 tsnlndvrtt cdtlttrht dltlnntlv nrlrdsislr mqqdmmskl dtevanlsv
 421 meemklvds k hgqliknft lqpppgprg kgdrgsqgpp gptgnkgqkg ekgepgpppg
 481 agergtigpv gppgergskg skgsqgpkgs rgspgkpgpq gpsgdpgppg
 ppgkdglpgp
 541 qpppgfqglq gtvgepgvpg prglpglpgv pgmpgpkgpp gppgpgsgame
 20 plalqnept
 601 asevnqcp h wknftdkcy f slekeifed aklfcedkss hlrvinsree qqwikkhtvg
 661 reshwigltd seqesewkw l dgspvdyknw kagqpdnwgs ghgpgedcag li-
 yagqwndf
 721 qcdeinnfic ekereavpss il
 25

SEQ ID NO: 45
 AAM34742 46-kDa collectin precursor [Bos taurus]
 gi|21105685|gb|AAM34742.1|AF509589_1|[21105685]
 30 sig_peptide 1..20
 Region 67..245 /region_name="collagen-like region"
 Region 245..371 /region_name="carbohydrate recognition domain"
 CDS 1..371 /gene="CL-46" /coded_by="join(AF509589.1:1454..1652,
 AF509589.1:5950..6066,AF509589.1:6402..6509,
 35 AF509589.1:6823..6930,AF509589.1:7289..7405,
 AF509589.1:8021..8104,AF509589.1:10318..10700)"
 ORIGIN 1 millplsvll lltqpwrsig aemkiysqkt langctlvvc rpeggplgr dgqdgregpq
 61 gekgdpgspg pagragrpgp agpigpkgn gsagepgpkg dtgppgppgm ppgagregps
 121 gkqgsmgppg tpgpkgdtp kggmgapgm gspgpaglk ergapgelga pgsagvagpa
 40 181 gaigpqgpg argppglkgd rgdpgergak gesgladvna lkqrvtileg qlqlqnafs
 241 rykkavlfpd qavgkkfk tagavksysd aqlcreakg qlasprsaee neavaqlvra
 301 knndaflsmn distegkfty ptgeslvysn wasgepnenn agqpencvqi yregkwndvp
 361 csepllvce f

45 SEQ ID NO: 47
 XP_139613 similar to collectin sub-family member 10; collectin liver 1; collectin 34
 [Mus musculus]
 gi|20903807|ref|XP_139613.1|[20903807]
 50 FEATURES Location/Qualifiers source 1..420 /organism="Mus musculus"
 /strain="C57BL/6J" /db_xref="taxon:10090" /chromosome="15"
 Protein 1..420 /product="similar to collectin sub-family member 10; collectin liver 1;
 collectin 34"

Region 152..269 /region_name="C-type lectin (CTL) or carbohydrate-recognition domain (CRD)" /note="CLECT" /db_xref="CDD:smart00034"
 Region 165..269 /region_name="Lectin C-type domain" /note="lectin_c" /db_xref="CDD:pfam00059"

5 Region 362..419 /region_name="Ubiquitin-conjugating enzyme E2, catalytic domain homologues" /note="UBCc" /db_xref="CDD:smart00212"
 Region 363..419 /region_name="Ubiquitin-conjugating enzyme" /note="UQ_con" /db_xref="CDD:pfam00179" CDS 1..420 /gene="LOC239447" /coded_by="XM_139613.1:1..1263" /db_xref="InterimID:239447"

10 ORIGIN 1 mngfrvllrs nlsmlillal lhfqslgldv dsrsaaevca thtispgpkg dddgerdgtge
 61 egkdgkvgrq gpkgvkgelg dmgaqgnigk sgpigkkgdk gekglgipg ekgkagticd
 121 cgryrkvvqg ldisvarlkt smkfiknvia gireteekfy yivqeenknyr eslthcring
 181 gmlampkdev vntliadyva ksgffrvfig vndleregqy vftdntplqn ysnwkeeps
 241 dpsghedcve mlssgrwndt echltmyfvs slqedliedc lreqglivqv tpanqellfg
 15 301 idtflgpmisc vyqrtgkqk lysqcrldwg lakkqtneta niatfckgae pnrgrpcgq
 361 kqemmtlmms gnkgittfpe sdnlfkwvgt mlgaagtide dlkyklslns pvvtliihpq

SEQ ID NO: 48
 XP_123211 similar to collectin sub-family member 12 [Mus musculus]
 20 gi|20876566|ref|XP_123211.1|[20876566]
 FEATURES Location/Qualifiers source 1..742 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /chromosome="18"
 Protein 1..742 /product="similar to collectin sub-family member 12"
 25 Region 79..320 /region_name="V-type ATPase 116kDa subunit family" /note="V_ATPase_sub_a" /db_xref="CDD:pfam01496"
 Region 92..337 /region_name="Intermediate filament protein" /note="filament" /db_xref="CDD:pfam00038" Region 607..731 /region_name="C-type lectin (CTL) or carbohydrate-recognition domain (CRD)" /note="CLECT" /db_xref="CDD:smart00034"

30 Region 624..732 /region_name="Lectin C-type domain" /note="lectin_c" /db_xref="CDD:pfam00059"
 CDS 1..742 /gene="LOC225157" /coded_by="XM_123211.1:77..2305" /db_xref="InterimID:225157"

35 ORIGIN 1 mkddfaeaaa vqsfgykrfg iqegtqctkc knnwalkfsi vlyilcall titvailgyk
 61 vvekmdnvd gmetshqtyd nklavesdl kklgdqagkk alstnselst frsdildlrq
 121 qlqektks knkdtleklq angdslvdrq sqlketlqnn sflittvnkt lqayngyvt
 181 lqqdtsvlqg nlqsmysqs vvimnlnln ltqvqqnli snlqqsvddt slaiqriknd
 241 fnlqqvflq akkdtwike kvqslqlaa nnsalakann dtledmnsq ssftgqmdni
 301 ttisqaneqs lkdldlhkd tenrtavkfs qleerfqvfe tdivniisni sytahhlrl
 40 361 tsnlndvrtt ctdlttrhtd dltslnntlv nirdsislr mqqdmmskl dtevanlsvv
 421 meemklvdsk hgqliknfti lqpppgprgp kgdrsgqgpp gptgnkgqkg ekgepgppgp
 481 agerdtigpv gppgergskg skgsqgpkgs rgspgkpgpq gpsgdpgppg
 ppgkdglpgp
 541 qgppgfqglq gtvgepgvpg prglpglpgv pgmpgpkgpp gppgpgame plalqneptp
 45 601 asevnqcpvh wknftdkcyf fslekeifed aklfcedkss hlvinfree qqwikhtvg
 661 reshwigltd seqesewkw dgsppvdyknw kagqpdnwgs ghgpgedcag li-
 yagqwndf
 721 qcdeinnfic ekereavpss il

50 SEQ ID NO: 49
 NP_571645 mannose binding-like lectin [Danio rerio]
 gi|18858997|ref|NP_571645.1|[18858997]
 sig_peptide 1..23

24

mat_peptide 24..251 /product="mannose binding-like lectin"
 Region 24..36 /region_name="N-terminal segment"
 Region 33..70 /region_name="Collagen triple helix repeat (20 copies)"
 /note="Collagen" /db_xref="CDD:pfam01391"
 5 Region 33..70 /region_name="Collagen triple helix repeat (20 copies)"
 /note="Collagen" /db_xref="CDD:pfam01391"
 Region 37..101 /region_name="collagen-like structure"
 Region 37..70 /region_name="Collagen triple helix repeat (20 copies)"
 /note="Collagen" /db_xref="CDD:pfam01391"
 10 Region 71..74 /region_name="break in collagen structure"
 Region 102..132 /region_name="neck region"
 Region 133..251 /region_name="carbohydrate recognition domain" /note="CRD"
 Region 134..247 /region_name="C-type lectin (CTL) or carbohydrate-recognition
 domain (CRD)" /note="CLECT" /db_xref="CDD:smart00034"
 15 Region 146..247 /region_name="Lectin C-type domain" /note="lectin_c"
 /db_xref="CDD:pfam00059"
 CDS 1..251 /gene="mbi" /coded_by="NM_131570.1:68..823" /note="collectin with
 structural homology to mannose-binding lectin but with a predicted carbohydrate
 specificity for galactose;mannose binding-like lectin" /db_xref="LocusID:58091"
 20 ORIGIN 1 mallklflga lllqlvlql magaadpql ncpayagvpg tpghnglpgr dgrvgrdgan
 61 gpkgekgpgg vnvqgppgka gppgpagakg ergpsglpgq dcmsdskse lqlsdskial
 121 iekvvnfktf kkvgqkyvt ddveetfdkg mqycssngga lvprtleem allkvfvssa
 181 fklrfirtd rekegefvdtd drklftnw gpnqpdnykg aqdcgaiads glwddvscds
 241 lypiiceiei k

25

SEQ ID NO: 50

NP_569057 collectin sub-family member 12, isoform I; scavenger receptor with C-type lectin; collectin placenta 1 [Homo sapiens]
gi|18641360|ref|NP_569057.1|18641360]

5

FEATURES Location/Qualifiers source 1..742 /organism="Homo sapiens"
/db_xref="taxon:9606" /chromosome="18" /map="18pter-p11.3"
Protein 1..742 /product="collectin sub-family member 12, isoform I" /note="isoform I
is encoded by transcript variant I; scavenger receptor with C-type lectin; collectin
placenta 1"

10

Region 79..328 /region_name="V-type ATPase 116kDa subunit family"
/note="V_ATPase_sub_a" /db_xref="CDD:pfam01496" Region 443..589
/region_name="collagen-like domain"

15

Region 607..731 /region_name="C-type lectin (CTL) or carbohydrate-recognition
domain (CRD)" /note="CLECT" /db_xref="CDD:smart00034"

Region 624..732 /region_name="Lectin C-type domain" /note="lectin_c"
/db_xref="CDD:pfam00059"

Region 668..719 /region_name="Beta-lactamase" /note="beta-lactamase"
/db_xref="CDD:pfam00144"

20

CDS 1..742 /gene="COLEC12" /coded_by="NM_130386.1:172..2400"
/db_xref="LocusID:81035"

ORIGIN 1 mkddfaeaaa vqsfgykrfg iqegtqctkc knnwalkfsi illyilcall titvailgyk

25

61 vvekmdnvtg gmetsrqtyd dklavesdl kklgdqtgkk aistnselst frsdildlrq

121 qlreitekts knkdtleklq asgdalvdrq sqlketlenn sflittvnkt lqayngyvtn

181 lqqdtsvlqg nlqnqmyshn vvimnlnln ltqvqqnli tnlqrsvddt sqaiqriknd

241 fqnlqqvflq akkdtldwike kvqslqltaa nnsalakann dtledmnsqf nsftgqmeni

301 ttisqaneqn lkdldlhkd aenrtalkfn qleerfqlfe tdvniisni sytahhlrtf

361 tsnlnevrtt cdtltkhtd dltslntla nirdsvslr mqqdlmrsrl dtevanlsvi

421 meemklvdsk hgqliknfti lqgppgprgp rgdrqsqgpp gptgnkgqkg ekgepgppgp

30

481 agergpiga gppgerggkg skgsqgpkgs rgspgkpgpq gpsgdpgppg

ppgkeglpgp

541 qgppgfqglq gtvgepgvpg prglpglpgv pgmpgpkgpp gppgpgsavv plalqneptp

601 apedngcpqh wknftdkcyy fsvekeifed akfcedkss hlvfintree qqwikkmvg

661 reshwiglt dserenewkwf dgtsdyknw kagqpdnwh ghgpgedcag liyagqwndf

35

721 qcetvnnfic ekdretvlss al

SEQ ID NO: 51

NP_110408 collectin sub-family member 12, isoform II; scavenger receptor with C-type lectin; collectin placenta 1 [Homo sapiens]

40

gi|18641358|ref|NP_110408.2|18641358]

FEATURES Location/Qualifiers source 1..622 /organism="Homo sapiens"

/db_xref="taxon:9606" /chromosome="18" /map="18pter-p11.3"

Protein 1..622 /product="collectin sub-family member 12, isoform II" /note="isoform
II is encoded by transcript variant II; scavenger receptor with C-type lectin; collectin
placenta 1"

45

Region 79..328 /region_name="V-type ATPase 116kDa subunit family"
/note="V_ATPase_sub_a" /db_xref="CDD:pfam01496"

Region 443..589 /region_name="collagen-like domain"

CDS 1..622 /gene="COLEC12" /coded_by="NM_030781.2:172..2040"

50

/db_xref="LocusID:81035"

ORIGIN 1 mkddfaeaaa vqsfgykrfg iqegtqctkc knnwalkfsi illyilcall titvailgyk

61 vvekmdnvtg gmetsrqtyd dklavesdl kklgdqtgkk aistnselst frsdildlrq

121 qlreitekts knkdtleklq asgdalvdrq sqlketlenn sflittvnkt lqayngyvtn

181 lqqdtsvlqg nlqnqmyshn vvimnlnln ltqvqqnli tnlqrsvddt sqaiqriknd
 241 fqnllqqvflq akkdtldwike kvqslqtlaa nnsalakann dtledmnsq nsftgqmeni
 301 ttisqaneqn lkdldlhkd aenrtakfn qleerfqlfe tdivniisni sytahhlrt
 361 tsnlnevrtt ctdlttkhtd dltslnntla nirdsvslr mqqdlmrsrl dtevarisvi
 5 421 meemklvdsk hgqliknfti lqpppgprgp rgdrsgqgpp gptgnkgqkg ekgepgppgp
 481 agergpigpa gppgerggkg skgsqgpkgs rgspgkpgpq gpsgdpgppg
 ppgkeglpgp
 541 qgppgfqglq gtvgepgvpg prglpglpgv pgmpgpkgpp gppgpsgavv plalqnept
 601 apednsksk slqpggqgsa ca

10

SEQ ID NO: 52

NP_569716 collectin sub-family member 12 [Mus musculus]

gi|18485494|ref|NP_569716.1|[18485494]

FEATURES Location/Qualifiers source 1..742 /organism="Mus musculus"

15

/db_xref="taxon:10090"

Protein 1..742 /product="collectin sub-family member 12"

Region 79..320 /region_name="V-type ATPase 116kDa subunit family"

/note="V_ATPase_sub_a" /db_xref="CDD:pfam01496"

20

Region 607..731 /region_name="C-type lectin (CTL) or carbohydrate-recognition

domain (CRD)" /note="CLECT" /db_xref="CDD:smart00034"

Region 629..732 /region_name="Lectin C-type domain" /note="lectin_c"

/db_xref="CDD:pfam00059"

CDS 1..742 /gene="Colec12" /coded_by="NM_130449.1:77..2305"

/db_xref="LocusID:140792" /db_xref="MGD:2152907"

25

ORIGIN 1 mkddfaeeee vsfgykrfg ihegtqctkc innwalkfsi vlyilcall titvailgyk

61 vvekmdnvsd gmetshqtyd nklitavesdl kklgdqagkk alstnselst frsdildlrq

121 qlqeitekts knkdtleklq angdsvdrq sqlketlqnn sflittvnkt lqayngyvt

181 lqqdtnvlqg nlqsqmysqs vvimnlnln ltqvqqnli snlqsvddt slaiqriknd

30

241 fqnllqqvflq akkdtldwike kvqslqtlaa nnsalakann dtledmnsq ssftgqmdni

301 ttisqaneqs lkdldlhkd tenrtavkfs qleerfqvfe tdivniisni sytahhlrt

361 tsnlndvwt ctdlttrhtd dltslnntlv nirdsvslr mqqdmmrskl dtevarisv

421 meemklvdsk hgqliknfti lqpppgprgp kgdrsgqgpp gptgnkgqkg ekgepgppgp

481 agertigpv gppgergskg skgsqgpkgs rgspgkpgpq gpsgdpgppg

ppgkdglpgp

35

541 qgppgfqglq gtvgepgvpg prglpglpgv pgmpgpkgpp gppgpsgame plalqnept

601 asevnqcpvh wknftdkcy fslekeiled aklfcedkss hlvfinsree qqwikhtvg

661 reshwiglt seqesewkw dgspvdyknw kagqpdnws ghgpgedcag li-

yagqwndf

721 qcdeinnfic ekereavpss il

40

SEQ ID NO: 53

AAL61856 43kDa collectin precursor [Bos taurus]

gi|18252111|gb|AAL61856.1|[18252111]

45

FEATURES Location/Qualifiers source 1..321 /organism="Bos taurus"

/db_xref="taxon:9913"

Protein 1..321 /product="43kDa collectin precursor" /name="CL-43; conglutinin; SP-D"

Region 1..166 /region_name="collagen-like"

50

Region 167..193 /region_name="alpha-helical neck"

Region 195..321 /region_name="carbohydrate-recognition domain"

28

Region 126..236 /region_name="C-type lectin (CTL) or carbohydrate-recognition domain (CRD)" /note="CLECT" /db_xref="CDD:smart00034"
 Region 135..237 /region_name="Lectin C-type domain" /note="lectin_c" /db_xref="CDD:pfam00059"
 5 CDS 1..239 /gene="Mbl1" /coded_by="NM_010775.1:121..840" /db_xref="LocusID:17194" /db_xref="MGD:96923"
 ORIGIN 1 mlllpilpvl lcvsvsssg sqtcedtlkt csviacgrdg rdgpkgekge pgqglrglqg
 61 ppgklgppgs vsgpgspgpk gqkgdhgdnr aieeklanme aeirilkskl qltnklhafs
 121 mgkksqgklf vtnhekmpfs kvkslctelq gtvaiprae enkaieqvat giaflgitde
 10 181 ategqfmyvt ggrltylnwk kdepnnhgsg edcvilndng lwndiscqas fkavcefpā

SEQ ID NO: 55
 NP_034906 mannose binding lectin, serum (C) [Mus musculus]
 gi|6754656|ref|NP_034906.1|[6754656]
 15 sig_peptide 1..18
 Region 120..241 /region_name="C-type lectin (CTL) or carbohydrate-recognition domain (CRD)" /note="CLECT" /db_xref="CDD:smart00034"
 Region 140..242 /region_name="Lectin C-type domain" /note="lectin_c" /db_xref="CDD:pfam00059"
 20 CDS 1..244 /gene="Mbl2" /coded_by="NM_010776.1:177..911"
 /note="polysaccharide-binding component of RaRF; sequence similarity to man-
 nose-binding proteins" /db_xref="LocusID:17195" /db_xref="MGD:96924" ORIGIN 1
 misiftsfill cvvtvyaet ltegvqnsdp vvtcsspgln gfpkgdgrdg akgekgepgg
 61 glrglqgppg kvgptgppgn pglkgavgpk gdrgrdraefd tseidseiaa lrselralrn
 25 121 wwlfslsekv gkkyfvssvk kmsldrvkal csefqgsvat prnaeensa qkvakdiayl
 181 gitdvrvs fedltgnrvr ytnwndgepn ntgdgedcvv ilgngkwndv pcsdsflaic
 241 efsd

SEQ ID NO: 56
 30 NP_006429 collectin sub-family member 10; collectin liver 1; collectin 34 [Homo sapiens] gi|5453619|ref|NP_006429.1|[5453619]
 FEATURES Location/Qualifiers source 1..277 /organism="Homo sapiens"
 /db_xref="taxon:9606" /chromosome="8" /map="8q23-q24.1"
 Protein 1..277 /product="collectin sub-family member 10" /note="collectin liver 1;
 35 collectin 34"
 Region 152..271 /region_name="C-type lectin (CTL) or carbohydrate-recognition domain (CRD)" /note="CLECT" /db_xref="CDD:smart00034"
 Region 165..272 /region_name="Lectin C-type domain" /note="lectin_c" /db_xref="CDD:pfam00059"
 40 CDS 1..277 /gene="COLEC10" /coded_by="NM_006438.2:76..909" /db_xref="LocusID:10584"
 ORIGIN 1 mngfasllr nqfilvlfl lqiqlglldi dsrptaevca thtispgpkg ddgekdpge
 61 egkhgkvgrm gpkgikgelg dmgrdnigk tpgigkkgdk gekgllgipg ekgkagtvc
 121 cgryrkfvqg ldisiarlkt smkfvknvia gireteekfy yivqeeknyr eslthcrirg
 45 181 gmlampkdea antliadyva ksgffrvfig vndleregqy mftdntplqn ysnwnegeps
 241 dpyghedcve mlssgrwndt echltmyfvc efikkkk

SEQ ID NO: 57
 BAB72147 collectin placenta 1 [Homo sapiens]
 50 gi|17026101|dbj|BAB72147.1|[17026101]
 FEATURES Location/Qualifiers source 1..742 /organism="Homo sapiens"
 /db_xref="taxon:9606" /sex="female" /tissue_lib="placenta"
 Protein 1..742 /product="collectin placenta 1"

CDS 1..742 /gene="CL-P1" /coded_by="AB005145.1:71..2299"

ORIGIN 1 mkddfaeeree vqsfgykrfg iqegtqctkc knnwalkfsi illyilcall titvailgyk
 61 vvekmdnvtg gmetsrqtyd dkltavesdl kklgdqtkk aistnselst frsdildlrq
 5 121 qlreitekts knkdtleklq asgdalvdrq sqlketlenn sflittvnkt lqayngyvtn
 181 lqqdtsvlqg nlqnqmyshn vvimnlnln ltqvqqmli tnlqrsvddt sqaiqriknd
 241 fnlqqvflq akkdtldwke kvqslqtlaa nnsalakann dtiedmnsq nsftgqmeni
 301 ttisqaneqn lkdldlhkd aenrtakfn qleerqlfe tdvniisni sytahhlrt
 361 tsnlnevrtt cdtltkhtd dltslnntla nirdsvslr mqqdlnsrldtevanlsvi
 10 421 meemklvdsk hgqliknfti lqppgprgp rgdrsgqgpp gptgnkgqkg ekgepgppgp
 481 agergpigpa gppgerggkg skgsqgpkgs rgspgkpgpq gpsgdpggp ppgkeglpgp
 541 qgppgfqglq gtvgepgvpg prglpglpgv pgmpgpkgpp gppgpgsavv plalqneptp
 601 apedngcpvh wknftdkcyy fsvekeifed aklfcedkss hlvfintree qqwikkmvg
 661 reshwigltd serenewkw dgtspdyknw kagqpdnwh ghgpgedcag liyagqwndf
 15 721 qcetvnnfic ekdretvlss al

SEQ ID NO: 58

AAF63470 mannose binding-like lectin precursor [Carassius auratus]
 gi|7542474|gb|AAF63470.1|AF227739_1[7542474]
 20 sig_peptide <1..13
 Region 14..25 /region_name="N-terminal segment"
 Region 26..93 /region_name="collagen-like structure"
 Region 60..63 /region_name="break in collagen structure"
 Region 94..124 /region_name="neck region" Region 125..246
 25 /region_name="carbohydrate recognition domain" /note="CRD"
 CDS 1..246 /gene="MBL" /coded_by="AF227739.1:<1..742" /note="collectin with
 structural homology to mannose-binding lectin but with a predicted carbohydrate
 specificity for galactose"

ORIGIN 1 lllqfalql ldgaepqnln cpayggvpgt pghnglpgrd grdgkdgaig pkgekgesgv
 61 svqgppgkag ppgtagckge rgpsgpggsp gsesvleslk seiqqkaki atfekvssvc
 121 hfrkvqkyy itdgvgnfd qglksmefg gtmvsprtsa enqallkvv ssglsgskpy
 181 igvtdrkteg qfvdteqkl tftnwpgqp ddykglqdcg viedtglwdd ggcgdirpim
 241 ceidik

SEQ ID NO: 59

AAF63469 mannose binding-like lectin precursor [Danio rerio]
 gi|7542472|gb|AAF63469.1|AF227738_1[7542472]
 sig_peptide 1..23
 40 mat_peptide 24..251 /product="mannose binding-like lectin"
 Region 24..36 /region_name="N-terminal segment"
 Region 37..101 /region_name="collagen-like structure"
 Region 71..74 /region_name="break in collagen structure"
 Region 102..132 /region_name="neck region"
 45 Region 133..251 /region_name="carbohydrate recognition domain" /note="CRD"
 CDS 1..251 /gene="mbi" /coded_by="AF227738.1:68..823" /note="collectin with
 structural homology to mannose-binding lectin but with a predicted carbohydrate
 specificity for galactose"

ORIGIN 1 maliklflga lllqlvlql magaadpqs ncpayagvpg tpghnglpgr dgrvgrdan
 61 gpkgekgepg vnvqppgka gppgpagak ergpsglpgq dcmsdsikse lqlsdlkial
 121 iekvvnfktf kkvqkyyvt ddveetfdkg mqycssnga lvprtle n allkvfssa
 181 fkrfritd rekegefvdt drklftnw gpnqpdnykg aqdcgaiads glwddvsdcs
 241 lypiceiei k

SEQ ID NO: 60

AAF63468 mannose binding-like lectin precursor [Cyprinus carpio]

gi|7542470|gb|AAF63468.1|AF227737_1[7542470]

5 sig_peptide 1..23
 mat_peptide 24..256 /product="mannose binding-like lectin"
 Region 24..35 /region_name="N-terminal segment"
 Region 36..103 /region_name="collagen-like structure"
 10 Region 70..73 /region_name="break in collagen structure"
 Region 104..134 /region_name="neck region"
 Region 135..256 /region_name="carbohydrate recognition domain" /note="CRD"
 CDS 1..256 /gene="MBL" /coded_by="AF227737.1:67..837" /note="collectin with
 structural homology to mannose-binding lectin but with a predicted carbohydrate
 15 specificity for galactose"
 ORIGIN 1 malfkflgt lllqfalql ldgaepqnl cpayggvpgt pghnglpgrd grdgkdgaig
 61 pkgekgesgv svqgppgkag ppgpagekge rgptgsqgsp gsesvleslk seiqqikaki
 121 atfekvasvg hfrqvgqky itdgvgtf d qglkfckdfg gtmvfprtsa enqallklvv
 181 ssglsskppy igvtdreteg rfvntegkql tftnwpggqp ddykglqdcg viedsglwd
 20 241 gscgdirpim ceidnk

SEQ ID NO: 61

AAK97540 surfactant protein A precursor [Gallus gallus]

gi|15420996|gb|AAK97540.1|AF411083_1[15420996]

25 sig_peptide 1..18
 Region 19..34 /region_name="N-terminal segment"
 Region 35..43 /region_name="putative collagen structure"
 Region 44..76 /region_name="putative coil structure"
 Region 77..97 /region_name="alpha-helical coil-coil structure; neck region"
 30 Region 98..222 /region_name="carbohydrate recognition domain"
 Site 121..123 /site_type="glycosylation"
 Site 181..183 /site_type="glycosylation" /note="conserved"
 CDS 1..222 /gene="SP-A" /coded_by="AF411083.1:61..729"
 ORIGIN 1 mlsysfcmia aavalltpch aqncagapel psipgvsgll glgalkryfg sllwpygeek
 35 61 lpecqwlqrq qdlstssdde lgnvlnlrq rilqlegvla ldgkitkvge kifasngkev
 121 nfssalesce etgglatprn neeenkaimg ivkqynryay lgikesdtag qfkyvnnqpl
 181 nytswqqyep ngkgtckcve mytdgnwkdr kcnlyrltvc ey

SEQ ID NO: 62

40 JN0450 conglutinin precursor – bovine gi|346501|pir||JN0450[346501]

FEATURES Location/Qualifiers source 1..371 /organism="Bos taurus"
 /db_xref="taxon:9913"

Protein 1..371 /product="conglutinin precursor" /note="C3b-binding protein"
 45 Region 1..20 /region_name="domain" /note="signal sequence"
 Region 21..371 /region_name="product" /note="conglutinin"
 Region 46..214 /region_name="region" /note="collagen-like"
 Site 63 /site_type="binding" /note="carbohydrate (Lys) (covalent)"
 Site 63 /site_type="modified" /note="5-hydroxylysine (Lys)"
 50 Region 75..371 /region_name="product" /note="conglutinin-N"
 Site 78 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 87 /site_type="binding" /note="carbohydrate (Lys) (covalent)"
 Site 87 /site_type="modified" /note="5-hydroxylysine (Lys)"

Site 96 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 99 /site_type="binding" /note="carbohydrate (Lys) (covalent)"
 Site 99 /site_type="modified" /note="5-hydroxylysine (Lys)"
 Site 108 /site_type="modified" /note="4-hydroxyproline (Pro)"
 5 Site 111 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 129 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 132 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 135 /site_type="binding" /note="carbohydrate (Lys) (covalent)"
 Site 135 /site_type="modified" /note="5-hydroxylysine (Lys)"
 10 Site 141 /site_type="binding" /note="carbohydrate (Lys) (covalent)"
 Site 141 /site_type="modified" /note="5-hydroxylysine (Lys)"
 Site 147 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 153 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 159 /site_type="binding" /note="carbohydrate (Lys) (covalent)"
 15 Site 159 /site_type="modified" /note="5-hydroxylysine (Lys)"
 Site 162 /site_type="binding" /note="carbohydrate (Lys) (covalent)"
 Site 162 /site_type="modified" /note="5-hydroxylysine (Lys)"
 Site 171 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 195 /site_type="modified" /note="4-hydroxyproline (Pro)"
 20 Site 198 /site_type="binding" /note="carbohydrate (Lys) (covalent)"
 Site 198 /site_type="modified" /note="5-hydroxylysine (Lys)"
 Site 210 /site_type="binding" /note="carbohydrate (Lys) (covalent)"
 Site 210 /site_type="modified" /note="5-hydroxylysine (Lys)"
 Region 248..369 /region_name="domain" /note="C-type lectin homology #label
 25 LCH"
 Site 337 /site_type="binding" /note="carbohydrate (Asn) (covalent)"

ORIGIN 1 mlllplsvll lltqpwrsig aemtftsqki lanactivmc splesglpgh dgqdgrecph
 61 gekgdpgspg pagragrpgw vgpigpkgn gfvgepgpkg dtgprgppgm pgpagregps
 30 121 gkqgsmgppg tpgpkgetgp kggvgapgiq gfpgpsglkg ekgapgetga pgragvtgps
 181 gaigpqgpgs argppglkgd rgdpgetgak gesglaevna lkqrvtildg hlrrfqnafs
 241 qykkavlfpd gqavgekifk tagavksysd aeqlcreakg qlasprssae neavtqmvr
 301 qeknaylsmn distegrfty ptgeilvyn wadgpnnsd egqpencvei fpdgkwndvp
 361 cskqllvice f

35

SEQ ID NO: 63

A57250 mannan-binding protein - chicken (fragment)

gi|1362725|pir||A57250[1362725]

40 FEATURES Location/Qualifiers source 1..30 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 Protein 1..30 /product="mannan-binding protein" /note="collectin"
 Site 28 /site_type="modified" /note="4-hydroxyproline (Pro)"
 ORIGIN 1 lltcdkpeek myscpiiqcs apavnglpqd

45

SEQ ID NO: 64

A53570 collectin-43 - bovine gi|1083017|pir||A53570[1083017]

50 FEATURES Location/Qualifiers source 1..301 /organism="Bos taurus"
 /db_xref="taxon:9913"
 Protein 1..301 /product="collectin-43" /note="lectin CL-43"
 Region 177..299 /region_name="domain" /note="C-type lectin homology #label
 LCH"

ORIGIN 1 eemdvysekt ltdpctlvvc appadslrgh dgrdgkegpq gekgdpग्pgpg mpgpagregp
61 sgrqgsmgpp gtpgpkgepg peggvgapgm pgspgpaglk gergapggpg
aigpqgpgsa

5 121 mgppglkgdr gdpgekgarg etsvlevdtl rqrmmlege vqlqnivtq yrkavlfpg
181 qavgekifkt agavksysda eqlcreakgq lasprssaen eavtqlvrak nkaylsmnd
241 iskegkftyp tggslidysnw apgepnnrak degpenclei ysdgnwndie creerlvce
301 f

SEQ ID NO: 65

10 AAF28384 lung surfactant protein A [Sus scrofa]
gi|6782434|gb|AAF28384.1|AF133668_1[6782434]

FEATURES Location/Qualifiers source 1..116 /organism="Sus scrofa"
/db_xref="taxon:9823"

15 Protein <1..116 /product="lung surfactant protein A" /function="involved in the innate
immune system and lipid homeostasis within the lung" /name="collectin; SPA; SP-A"
CDS 1..116 /gene="SFTPA" /coded_by="AF133668.1:<1..353"

ORIGIN 1 avgekvfstn gqsvafdvir elcaraggri aaprspeene aiasivkkhn tyaylgiveg
61 ptagdflyd gtpvnytnwy pgeprgrgke kcvemytdgq wndrncqqr laicef

20

SEQ ID NO: 66

AAF22145 lung surfactant protein D precursor; SPD; SP-D; CP4 [Sus scrofa]
gi|6760482|gb|AAF22145.2|AF132496_1[6760482]

25

sig_peptide 1..20

mat_peptide 21..378 /product="lung surfactant protein D"

CDS 1..378 /gene="SFTPD" /coded_by="AF132496.2:44..1180"

ORIGIN 1 mlllplsqli lltqpprslg aemktysqra vanacalvmc spmenglpgr dgrdgregpr
61 gekgdpplpg avragmpgl agpvgpkgdn gstgepgakg digpcgpppg

30

pgipgpagke

121 gpsgqqgnig ppgtpgpkge tgpkgevgal gmqgstgarg paglkgerga pgergap-
gsa

181 gaagpagatg pqgpgsargp pglkgdrpgp gergakgesg lpgitalrqv vetlqqqvqr

241 lqkafsqykk velfpngrgv gekifktggf ektfqdaqv ctqaggqmas prseteneal

35

301 sqlvtaqnka aflsmtdikt egnftyptge plvyawapq epnnnggssg aencveifpn

361 gkwndkacge lrivicef

SEQ ID NO: 67

40 P41317 MANNANOSE-BINDING PROTEIN C PRECURSOR (MBP-C) (MANNAN-
BINDING PROTEIN) (RA-REACTIVE FACTOR P28A SUBUNIT) (RARF/P28A)
gi|1346477|sp|P41317|MABC_MOUSE[1346477]

FEATURES Location/Qualifiers source 1..244 /organism="Mus musculus"
/db_xref="taxon:10090"

45

gene 1..244 /gene="MBL2"

Protein 1..244 /gene="MBL2" /product="MANNANOSE-BINDING PROTEIN C PRE-
CURSOR"

Region 1..18 /gene="MBL2" /region_name="Signal" /note="BY SIMILARITY."

Region 3 /gene="MBL2" /region_name="Conflict" /note="I -> L (IN REF. 1)."

50

Region 15 /gene="MBL2" /region_name="Conflict" /note="V -> A (IN REF. 1)."

Region 19..244 /gene="MBL2" /region_name="Mature chain" /note="MANNANOSE-
BINDING PROTEIN C."

- Bond bond(29) /gene="MBL2" /bond_type="disulfide" /note="INTERCHAIN (BY SIMILARITY)."
- Bond bond(34) /gene="MBL2" /bond_type="disulfide" /note="INTERCHAIN (BY SIMILARITY)."
- 5 Region 38..96 /gene="MBL2" /region_name="Domain" /note="COLLAGEN-LIKE (G-X-Y)."
- Site 43 /gene="MBL2" /site_type="hydroxylation" /note="(POTENTIAL)."
- Site 58 /gene="MBL2" /site_type="hydroxylation" /note="(POTENTIAL)."
- Site 69 /gene="MBL2" /site_type="hydroxylation" /note="(POTENTIAL)."
- 10 Site 78 /gene="MBL2" /site_type="hydroxylation" /note="(POTENTIAL)."
- Site 81 /gene="MBL2" /site_type="hydroxylation" /note="(POTENTIAL)."
- Region 149..242 /gene="MBL2" /region_name="Domain" /note="C-TYPE LECTIN (SHORT FORM)."
- Bond bond(151,240) /gene="MBL2" /bond_type="disulfide" /note="BY SIMILARITY."
- 15 Bond bond(218,232) /gene="MBL2" /bond_type="disulfide" /note="BY SIMILARITY."
- ORIGIN 1 msiftsfill cvwtvyaet ltegvqnsdp vvtcsspgln gfpkgdgrdg akgekgepgg
61 glrglqgppg kvqptgppgn pglkgavgpk gdrgrdraefd tseidseiaa lrselralrn
121 wwlfslsekv gkkyfvssvk kmsldrvkal csefqgsvat prnaeensa qkvakdiayl
20 181 gitdvrvs fedltgnrvr ytnwndgepn ntgdgedcvv ilngnkwndv pcsdsflaic
241 efsd
- SEQ ID NO: 68
- 25 P39039 MANNANOSE-BINDING PROTEIN A PRECURSOR (MBP-A) (MANNAN-BINDING PROTEIN) (RA-REACTIVE FACTOR POLYSACCHARIDE-BINDING COMPONENT P28B POLYPEPTIDE) (RARF P28B)
gi|729972|sp|P39039|MABA_MOUSE[729972]
- 30 FEATURES Location/Qualifiers source 1..239 /organism="Mus musculus"
/db_xref="taxon:10090"
gene 1..239 /gene="MBL1"
Protein 1..239 /gene="MBL1" /product="MANNANOSE-BINDING PROTEIN A PRE-CURSOR"
- 35 Region 1..17 /gene="MBL1" /region_name="Signal" /note="BY SIMILARITY."
Region 18..239 /gene="MBL1" /region_name="Mature chain" /note="MANNANOSE-BINDING PROTEIN A." Region 37..89 /gene="MBL1" /region_name="Domain" /note="COLLAGEN-LIKE (G-X-Y)."
- 40 Region 144..239 /gene="MBL1" /region_name="Domain" /note="C-TYPE LECTIN (SHORT FORM)."
- Bond bond(146,235) /gene="MBL1" /bond_type="disulfide" /note="BY SIMILARITY."
- Bond bond(213,227) /gene="MBL1" /bond_type="disulfide" /note="BY SIMILARITY."
- ORIGIN 1 mlllpllpvl lcwsvsssg sqtcedtlkt csviacgrdg rdgpkgekge pgqglrglqg
45 61 ppqklgppgs vgspspgpk gqkgdhgdnr aieeklanme aeirlikskl qltnklhafs
121 mgkksqgklf vtnhekmpfs kvkslctelq gtaiprae enkaieqvaf giaflgitde
181 ateqqfmyvt ggrltysnwk kdepnnhgsg edcvliidng lwndiscqas fkavcefp
- SEQ ID NO: 69
- 50 P42916 COLLECTIN-43 (CL-43) gi|1168967|sp|P42916|CL43_BOVIN[1168967]
FEATURES Location/Qualifiers source 1..301 /organism="Bos taurus"
/db_xref="taxon:9913"
Protein 1..301 /product="COLLECTIN-43"

Region 29..142 /region_name="Domain" /note="COLLAGEN-LIKE (G-X-Y)."
 Region 202..301 /region_name="Domain" /note="C-TYPE LECTIN (SHORT FORM)."
 Bond bond(204,299) /bond_type="disulfide" /note="BY SIMILARITY."
 5 Bond bond(277,291) /bond_type="disulfide" /note="BY SIMILARITY."
 ORIGIN 1 eemdvysekt ltdpctlvvc appadslrgh dgrdgkegpq gekgdpgppg mppgpagregp
 61 sgrqgsmgpp gtpgpkgepg peggvgapgm pgspgpaglk gergapgppg
 aigpqgppsga
 121 mgppglkgdr gdpgekgarg etsvlevdtl rqrnrnlege vqrlqnvitq yrkavlfpdg
 10 181 qavgekifkt agavksysda eqlcreakgq lasprssaen eavtqlvrak nkaylismnd
 241 iskegkftyp tggsl dysnw apgepgnrak degpenclei ysdgnwndie creerlvce
 301 f

.SEQ ID NO: 70
 15 CAB56155 DMBT1/8kb.2 protein [Homo sapiens]
 gi|5912464|emb|CAB56155.1|[5912464]
 sig_peptide 1..26
 mat_peptide 26..2412 /product="DMBT1/8kb.2 protein"
 CDS 1..2412 /gene="DMBT1" /coded_by="AJ243212.1:107..7345"
 20 /note="Sequence is an alternative splice form of the DMBT1 gene that is expressed
 in human adult trachea. Isoforms of DMBT1 are identical to the collectin binding
 protein gp-340. Full-length cDNA clone contains 1 bp deletions in codons 100 and
 1751, that were corrected by comparison with the genomic exons"
 ORIGIN 1 mgistvilem clwgqvlt ggwiprttdy aslipsevpl dttvaegspf pseltlelv
 25 61 aegspisles tlettvaegs lipsestles tvaegsdsgl alrlvngdgr cqrgrveilyr
 121 gswgavcdds wdtdanvvc rqlgcgwams apgnawfgqg sgpiatddvr csghe-
 sylws
 181 cphngwlshn cghgedagvi csaaqpqstl rpeswpvris ppvptegses slalrlvngg
 241 drcrgrvevl yrgswgtvcd dywdtdanv vcrqlgcgwa msapgnaqfg qsggpividd
 30 301 vrcsghesyl wscphngwlt hncghsedag vicsapqsrp tpspdtwpts hastagpess
 361 lalrlvnggd rcqgrvevly rgswwgtvcd swdtdanvv crqlgcgwat sapgnarfqq
 421 gsgpividdv rcsghesylw scphngwlsh ncqhsedagv icsaahswst pspdtlptit
 481 lpastvgses slalrlvngg drcqgrvevl yrgswgtvcd dswdtdanv vcrqlgcgwa
 541 mlapgnarfqq qsggpividd vrcsgnesyl wscphngwls hncghsedag vicsgpessl
 35 601 alrlvnggdr cqrgrvevlyr gswgtvcdds wdtdanvvc rqlgcgwams apgnarfqqg
 661 sgpiatddvr csghesylws cpnngwlshn cghhedagvi csaaqsrstp rpdltititl
 721 ppstvgsses lrlvngsd rcqgrvevly rgswwgtvcd swdtdanvv crqlgcgwat
 781 sapgnarfqq gsgpividdv rcsghesylw scphngwlsh ncghhedagv icsvsqsrpt
 841 pspdtwptsh astagpessl alrlvnggdr cqrgrvevlyr gswgtvcdds wdtdanvvc
 40 901 rqlgcgwats apgnarfqqg sgpiatddvr csgyesylws cphngwlshn cqhsedagvi
 961 csaaahswstp spdtlptitl pastvgsses lalrlvnggd rcqgrvevly qsgwtvcdd
 1021 swdtdanvv crqlgcgwam sapgnarfqq gsgpividda rcsghesylw scphngwlsh
 1081 ncghsedagv icsasqsrpt pspdtwptsh astagssessl alrlvnggdr cqrgrvevlyr
 1141 gswgtvcddy wdtdanvac rqlgcgwams apgnarfqqg sgpiatddvr csghesylws
 45 1201 cphngwlshn cghhedagvi csasqsqtp spdtwptsha stagssessl lrlvnggdrc
 1261 qgrvevlyrg swgtvcddyw dtdanvvcr qlgcgwatsa pgnarfqqgs gpividdvrc
 1321 sghesylwsc phngwlshnc ghhedagvic sasqsqtps pdtwptshas tagssessl
 1381 rlvggdrq grvevlyrgs wgtvcddywd tndanvvcrq lgcgwatsap gnarfqqsg
 1441 pividdvrcs ghesylwscph hngwlshncg hhedagvics afqsqtpsp dtwptsrast
 50 1501 agsestlrl lvggdrgrg rvevlyqgs wgtvcddywd tndanvvcrq lgcgwamsap
 1561 naqfgqgsgp ividdvrcsg hepylwscph ngwlshncg hhedagvicsa aqsqstprpd
 1621 twltltpal tvgsesslrl rlvggdrgr grvevlyrgs wgtvcddswd tndanvvcrq
 1681 lgcgwamsap gnarfqqsg pivlvdvrcs gnesylwscph hkgwlshncg hhedagvics

35

1741 atqinstttdd wwhpttttta rpsnccggfl fyasgtfssp sypayypnna kcvweievns
 1801 gyrinlgfsn lkleahhncs fdyveifdgs lnslllgki cndtrqifts synrmtihfr
 1861 sdisfqntgf lawynsfpsd atlrlnlns syglcagrve iyhggtwgav cddswtqiea
 1921 evvcrqlgcg ravsalignay fgsgsgpiti ddevcsgetes tlwqcrnrgw fshnchnred
 5 1981 agvicsgnhl stpafinit rpnnyscggl lsqpsgdfss pfypgnypnn akcvwdievq
 2041 nnyrvtvifr dvqleggcny dyievfdgpy rsspliarvc dgargsftss snfmsirfis
 2101 dhsitrrgrf aeyyspsnd stnlclpnh mqasvsrsyl qslgfsasdl vistwngyye
 2161 crpqitpnlv iftipysgcg tfkqadndti dysnlltaav sggiikrrtd lrihvscrml
 2221 qntwvdtmyi andtihvann tiqveevqyg nfdvnisfyt sssflypvtis rpyyvdlndq
 10 2281 lyvqaeilhs davltifvdt cvaspysndf tsitydlirs gcvrddtygp ysspslriar
 2341 frfrahfln rpsvylrck mvvcraydps srcyrgcvlr skrdvgsyqe kvdvvlpgiq
 2401 lqtpprree pr

SEQ ID NO: 71

15 BAA81747 collectin 34 [Homo sapiens] gi|5162875|dbj|BAA81747.1|[5162875]
 FEATURES Location/Qualifiers source 1..277 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 Protein 1..277 /product="collectin 34"
 CDS 1..277 /coded_by="AB002631.1:6..839"
 20 ORIGIN 1 mngfasllrr nqfillvfl lqiqlgldi dsrptaevca thtispgpkg ddekgdpge
 61 egkhgkvgrm gpkgikgelg dmngdrnigk tgpigkkgdk gekglgipg ekgkagtvc
 121 cgryrkfvqg ldisiarlkt smkfvknvia gireteekfy yivqeenyr eslthcrirg
 181 gmlampkdea antliadyva ksgffrvfig vndlereggy mftdntplqn ysnwnegeps
 241 dpyghedcve mlssgrwndt echltmyfvc efikkkk

25

SEQ ID NO: 72

AAB94071 mannan-binding lectin; collectin [Gallus gallus]
 gi|2736145|gb|AAB94071.1|[2736145]
 FEATURES Location/Qualifiers source 1..238 /organism="Gallus gallus"
 30 /strain="White Leghorn" /db_xref="taxon:9031" /tissue_type="liver"
 Protein 1..>238 /product="mannan-binding lectin" /name="c-type lectin"
 /note="mannan-binding protein; MBP; mannose-binding protein; MBL; collectin"
 CDS 1..238 /gene="cMBL" /coded_by="AF022226.1:1..>714"
 ORIGIN 1 mmatsllttd kpeekmyscp iicqapavn glpgrdgrdg pkgekdpge glrglqglpg
 35 61 kagpqglkge vgpqgekqk gergivvtd lhrqitdlea kirvleddls rykkalskd
 121 vvnigkkmfv stgkkynek gkslcakags vlasprneae ntalkdlidp ssqayigisd
 181 aqtegrfmyl sgpltytnw kpgpenhkn edcaviedsg kwndldcsns nifilcel

SEQ ID NO: 73

40 AAB36019 mannan-binding protein, MBP=lectin {N-terminal} [chickens, serum,
 Peptide Partial, 30 aa] [Gallus gallus] gi|1311692|gb|AAB36019.1|[1311692]
 FEATURES Location/Qualifiers source 1..30 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 Protein 1..30 /partial /product="mannan-binding protein" /name="lectin" /note="MBP"
 45 ORIGIN 1 lltcdkpeek myscliqcs apavnglpgd

SEQ ID NO: 74

AAB27504 conglutinin (N) {N-terminal} [cattle, Peptide Partial, 60 aa] [Bos taurus]
 gi|386660|gb|AAB27504.1|[386660]
 50 FEATURES Location/Qualifiers source 1..60 /organism="Bos taurus"
 /db_xref="taxon:9913"
 Protein 1..60 /partial /product="conglutinin (N)"
 ORIGIN 1 aemtffsqki lanactlvmc splsglpgd dgqdgrecph gekgdpqspg pagragrpgw

SEQ ID NO: 75

CAA53511 collectin-43 [Bos taurus] gi|499385|emb|CAA53511.1|[499385]

FEATURES Location/Qualifiers source 1..301 /organism="Bos taurus"

/db_xref="taxon:9913" /tissue_type="liver" /clone_lib="lambda gt 11"

Protein 1..301 /product="collectin-43"

mat_peptide 1..301 /product="collectin-43"

CDS 1..301 /coded_by="X75912.1:<1..906" /db_xref="SWISS-PROT:P42916"

ORIGIN 1 eemdvyxekt ltdpctlvvc appadslrgh dgrdgkegpq gekgdpqpgg mpgpagregg

61 sgrqgsmgpp gtpgpkgepg pegvggapgm pgspgpaglk gergapggg

aigpqgpgsa

121 mgppglkgdr gdpgekgarg etsvlevdtl rqrmmnlege vqrlqnivtq yrkavlfpgd

181 qavgekifkt agavksysda eqlcreakgq lasprssaen eavtqlvrak nkhaylsmnd

241 iskegkftyp tggsl dysnw apgepgnrak degpenclei ysdgnwndie creerlvce

301 f

SEQ ID NO: 76

AAA82010 mannose-binding protein C [Mus musculus]

gi|773288|gb|AAA82010.1|[773288]

FEATURES Location/Qualifiers source 1..244 /organism="Mus musculus"

/strain="BALB/c" /db_xref="taxon:10090" /clone="Lambda 14 and 52; Cos11A"

/clone_lib="NIH/3T3 Swiss mouse embryo cell line and BALB/c pWE15 cosmid library"

Protein 1..244 /product="mannose-binding protein C"

Site 1..59 /site_type="signal-peptide" /note="signal-peptide and collagen-like region"

mat_peptide <59..>98 /product="mannose-binding protein C" /note="collagen-like domain"

mat_peptide <98..>121 /product="mannose-binding protein C" /note="linking-peptide domain"

mat_peptide <121..244 /product="carbohydrate recognition domain"

CDS 1..244 /gene="Mbl2" /coded_by="join(U09013.1:470..644,U09014.1:43..159,U09015.1:97..165,U09016.1:576..949)"

ORIGIN 1 msiftsflll cvvtvyaet ltegvqncsp vvtcsspgln gfpkgdgrdg akgekgepgq

61 glrlgqpgg kvgptgppgn pglkgavgpk gdrdraefd tseidseiaa lrselraln

121 wvlfslsekv gkkyfvssvk kmsldrvkal csefqgsvat prnaeensa qkvakdiayl

181 gitdvrvegs fedltgnrvr ytnwndgepn ntgdgedcvv ilgngkwndv pcsdsflaic

241 efsd

SEQ ID NO: 77

AAA82009 mannose-binding protein A [Mus musculus]

gi|773280|gb|AAA82009.1|[773280]

sig_peptide 1..18

mat_peptide 19..239 /product="unnamed"

mat_peptide 19..>52 /product="mannose-binding protein A" /note="collagen-like region"

mat_peptide <52..>91 /product="mannose-binding protein A" /note="collagen-like domain"

mat_peptide <91..>116 /product="mannose-binding protein A" /note="linking-peptide domain"

CDS 1..239 /gene="Mbl1" /coded_by="join(U09007.1:275..428,U09008.1:287..403,U09009.1:166..240,U09010.1:78..451)"

ORIGIN 1 mlllpplpvl lcvsvsssg sqtcedtlkt csviacgrdg rdgpkgekge pgqglrlgq

61 ppgklgppgs vgspspgpk gqkgdhgdnr aieeklanme aeirilkskl qltnklhafs
 121 mgkksqgklf vtnhekmfks kvkslctelq gtvaiprnae enkaieqvaf giaflgitde
 181 ategqfmyvt ggrltywnwk kdepnnhgsg edcvilndng lwndiscqas fkavcefpaf

5

Lung surfactant protein

SEQ ID NO: 78

10 P35247 Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D)
 gi|464486|sp|P35247|PSPD_HUMAN[464486]

FEATURES Location/Qualifiers source 1..375 /organism="Homo sapiens"
 /db_xref="taxon:9606"

15 gene 1..375 /gene="SFTPD" /note="SFTP4; PSPD"
 Protein 1..375 /gene="SFTPD" /product="Pulmonary surfactant-associated protein D precursor"
 Region 1..20 /gene="SFTPD" /region_name="Signal" /note="BY SIMILARITY."
 Region 21..375 /gene="SFTPD" /region_name="Mature chain"
 20 /note="PULMONARY SURFACTANT-ASSOCIATED PROTEIN D."
 Region 31 /gene="SFTPD" /region_name="Conflict" /note="M -> T (IN REF. 2)."
 Region 46..222 /gene="SFTPD" /region_name="Domain" /note="COLLAGEN-LIKE."
 Region 59 /gene="SFTPD" /region_name="Conflict" /note="P -> F (IN REF. 3)."
 Site 78 /gene="SFTPD" /site_type="hydroxylation" /note="(BY SIMILARITY)."
 25 Site 87 /gene="SFTPD" /site_type="hydroxylation" /note="(BY SIMILARITY)."
 Site 90 /gene="SFTPD" /site_type="glycosylation" /note="N-LINKED (GLCNAC...) (POTENTIAL)."
 Site 96 /gene="SFTPD" /site_type="hydroxylation" /note="(BY SIMILARITY)."
 Site 99 /gene="SFTPD" /site_type="hydroxylation" /note="(BY SIMILARITY)."
 30 Region 122 /gene="SFTPD" /region_name="Conflict" /note="A -> P (IN REF. 2)."
 Site 171 /gene="SFTPD" /site_type="hydroxylation" /note="(BY SIMILARITY)."
 Site 177 /gene="SFTPD" /site_type="hydroxylation" /note="(BY SIMILARITY)."
 Region 180 /gene="SFTPD" /region_name="Conflict" /note="T -> A (IN REF. 2)."
 Region 206 /gene="SFTPD" /region_name="Conflict" /note="D -> P (IN REF. 3)."
 35 Region 223..252 /gene="SFTPD" /region_name="Domain" /note="COILED COIL (POTENTIAL)."
 Region 227..253 /gene="SFTPD" /region_name="Helical region"
 Region 254..256 /gene="SFTPD" /region_name="Hydrogen bonded turn"
 Region 257..260 /gene="SFTPD" /region_name="Beta-strand region"
 40 Region 261..262 /gene="SFTPD" /region_name="Hydrogen bonded turn"
 Region 263..272 /gene="SFTPD" /region_name="Beta-strand region"
 Region 274..283 /gene="SFTPD" /region_name="Helical region"
 Region 279..375 /gene="SFTPD" /region_name="Domain" /note="C-TYPE LECTIN (SHORT FORM)."
 45 Bond bond(281,373) /gene="SFTPD" /bond_type="disulfide"
 Region 284..285 /gene="SFTPD" /region_name="Hydrogen bonded turn"
 Region 287..288 /gene="SFTPD" /region_name="Beta-strand region"
 Region 294..307 /gene="SFTPD" /region_name="Helical region"
 Region 308 /gene="SFTPD" /region_name="Hydrogen bonded turn"
 50 Region 311..316 /gene="SFTPD" /region_name="Beta-strand region"
 Region 321..322 /gene="SFTPD" /region_name="Hydrogen bonded turn"
 Region 325 /gene="SFTPD" /region_name="Beta-strand region"
 Region 327..328 /gene="SFTPD" /region_name="Hydrogen bonded turn"

38

Region 331 /gene="SFTPD" /region_name="Beta-strand region"
 Region 337 /gene="SFTPD" /region_name="Beta-strand region"
 Region 339..340 /gene="SFTPD" /region_name="Hydrogen bonded turn"
 Region 345..347 /gene="SFTPD" /region_name="Helical region"
 5 Bond bond(351,365) /gene="SFTPD" /bond_type="disulfide"
 Region 351..354 /gene="SFTPD" /region_name="Beta-strand region"
 Region 356..357 /gene="SFTPD" /region_name="Hydrogen bonded turn"
 Region 360..363 /gene="SFTPD" /region_name="Beta-strand region"
 Region 365..366 /gene="SFTPD" /region_name="Hydrogen bonded turn"
 10 Region 369..375 /gene="SFTPD" /region_name="Beta-strand region"
 Region 374 /gene="SFTPD" /region_name="Conflict" /note="E -> EH (IN REF. 3)."
 ORIGIN 1 mlflfalslv lltqplgyle aemktyshrt mpsactlvmc ssvesglpgr dgrdgregpr
 61 gekgdpglpg aagqagmpgq agpvpgkgnv gsvgepgpkg dtgpsgppgp
 pgvpgpagre
 15 121 galgkqgnig pqgkpgpkge agpkgevgap gmqgsagarg lagpkgergv
 pgergvpgnt
 181 gaagsagamg pqgspgargp pglkgdkgip gdkgakgesg lpdvaslrqq vealqgqvqh
 241 lqaafsqqyk velfpngqsv gekifktagf vkpfteaql ctqaggqlas prsaaenaal
 301 qqlvvaknea aflsmtdskt egkftyptge slvysnwapg epnddggsed cveiftngkw
 20 361 ndraccgkrl vvcef

SEQ ID NO: 79

NP_002395 microfibrillar-associated protein 4; microfibril-associated glycoprotein 4
 [Homo sapiens] gi|23111005|ref|NP_002395.1||23111005]

25 FEATURES Location/Qualifiers source 1..255 /organism="Homo sapiens"
 /db_xref="taxon:9606" /chromosome="17" /map="17p11.2"
 Protein 1..255 /product="microfibrillar-associated protein 4" /note="microfibril-
 associated glycoprotein 4"
 30 Region 36..255 /region_name="smart00186, FBG, Fibrinogen-related domains
 (FReDs); Domain present at the C-termini of fibrinogen beta and gamma chains,
 and a variety of fibrinogen-related proteins, including tenascin and Drosophila
 scabrous"
 Region 38..254 /region_name="pfam00147, fibrinogen_C, Fibrinogen beta and
 35 gamma chains, C-terminal globular domain"
 CDS 1..255 /gene="MFAP4" /coded_by="NM_002404.1:26..793"
 /db_xref="LocusID:4239" /db_xref="MIM:600596"
 ORIGIN 1 mkallalpll llstppcap qvsgirgdal erfclqqpld cddiyaqgyq sdgvvlylps
 61 gpsvpvpvfc dmtteggkwt vfqkrfngsv sffrgwndyk lgfgradgey wglqnmhll
 40 121 tlkqkyelrv dledfennta yakyadfsis pnavsaeedg ytlfvagfed ggagdslyh
 181 sgqkfstfdr dqdlfvqnc a lssgafwfr schfanlngf ylggshlsya nginwaqwkq
 241 fyyslkrtem kirra

SEQ ID NO: 80

45 1KMRA Chain A, Solution Nmr Structure Of Surfactant Protein B (11-25) (Sp- B11-
 25). gi|22219056|pdb|1KMR|A|22219056]

50 FEATURES Location/Qualifiers source 1..15 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 SecStr 3..11 /sec_str_type="helix" /note="helix 1"
 ORIGIN 1 cralikriqa mipkg

SEQ ID NO: 81

P50404 Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D)
 gi|1709879|sp|P50404|PSPD_MOUSE[1709879]
 FEATURES Location/Qualifiers source 1..374 /organism="Mus musculus"
 /db_xref="taxon:10090"
 5 gene 1..374 /gene="SFTPD" /note="SFTP4"
 Protein 1..374 /gene="SFTPD" /product="Pulmonary surfactant-associated protein D precursor"
 Region 1..19 /gene="SFTPD" /region_name="Signal" /note="BY SIMILARITY."
 Region 20..374 /gene="SFTPD" /region_name="Mature chain"
 10 /note="PULMONARY SURFACTANT-ASSOCIATED PROTEIN D."
 Region 45..221 /gene="SFTPD" /region_name="Domain" /note="COLLAGEN-LIKE."
 Site 89 /gene="SFTPD" /site_type="glycosylation" /note="N-LINKED (GLCNAC...)
 (POTENTIAL)."
 Region 222..253 /gene="SFTPD" /region_name="Domain" /note="COILED COIL
 15 (POTENTIAL)."
 Region 278..374 /gene="SFTPD" /region_name="Domain" /note="C-TYPE LECTIN
 (SHORT FORM)."
 Bond bond(280,372) /gene="SFTPD" /bond_type="disulfide" /note="BY
 SIMILARITY."
 20 Bond bond(350,364) /gene="SFTPD" /bond_type="disulfide" /note="BY
 SIMILARITY."
 ORIGIN 1 mlpfslmvl lvqplgnlga emkslsqrsv pntctlvmsc ptenglprgd grdgregprg
 61 ekgdpglpqp mglsglqgpt gpvpgpkeng sagepgpkge rglspppglp gipgpagkeg
 121 psgkqgnigp qgkpgpkgea gpkgevgapq mqgstgakgs tgpkgergap
 25 gvqgapgnag
 181 aagpagpagp qgapgsrgpp glkgdrvgp drgikgesgl pdsaalrqm ealkgklqlr
 241 evafshyqka alfpdgrsvg dkifrtadse kpfedaqemc kqaggqlasp rsatenaaiq
 301 qlitahnkaa flsmtdivgte gkftyptgep lvysnwapge pnnnggaenc veiftnqwn
 361 dkacgeqlv icef
 30
 SEQ ID NO: 82
 P06908 Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
 (PSAP) gi|1172693|sp|P06908|PSPA_CANFA[1172693]
 35 FEATURES Location/Qualifiers source 1..248 /organism="Canis familiaris"
 /db_xref="taxon:9615"
 gene 1..248 /gene="SFTPA1" /note="SFTPA; SFTP1"
 Protein 1..248 /gene="SFTPA1" /product="Pulmonary surfactant-associated protein
 A precursor"
 40 Region 1..17 /gene="SFTPA1" /region_name="Signal"
 Region 18..248 /gene="SFTPA1" /region_name="Mature chain"
 /note="PULMONARY SURFACTANT-ASSOCIATED PROTEIN A."
 Site 20 /gene="SFTPA1" /site_type="glycosylation" /note="N-LINKED (GLCNAC...)
 (POTENTIAL)."
 45 Region 28..100 /gene="SFTPA1" /region_name="Domain" /note="COLLAGEN-
 LIKE."
 Region 153..248 /gene="SFTPA1" /region_name="Domain" /note="C-TYPE LECTIN
 (SHORT FORM)."
 Bond bond(155,246) /gene="SFTPA1" /bond_type="disulfide" /note="BY
 50 SIMILARITY."
 Site 207 /gene="SFTPA1" /site_type="glycosylation" /note="N-LINKED (GLCNAC...)
 (PROBABLE)."

Bond bond(224,238) /gene="SFTPA1" /bond_type="disulfide" /note="BY
SIMILARITY."

ORIGIN 1 mwlrlclal tllmvsgien ntkdvcvgnp gipgtpgshg lpgrdgrdgv kgdpgpppgpl
61 gppggmpgph gpngmtgapg vagergekge pgergppglp asldeelqtl lhdrlhqilq
121 tmglvslhes llvgrkvfs snaqsinfnd iqelcagagg qiaapmspee neavasivkk
181 yntyaylgiv espdsgdfqy mdgapvnytn wypgeprgrg keqcvemytd gqwnnknclq
241 yrlaicef

SEQ ID NO: 83

P12842 Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
(PSAP) gi|131413|sp|P12842|PSPA_RABIT[131413]

FEATURES Location/Qualifiers source 1..247 /organism="Oryctolagus cuniculus"
/db_xref="taxon:9986"

gene 1..247 /gene="SFTPA1" /note="SFTPA; SFTP1"

Protein 1..247 /gene="SFTPA1" /product="Pulmonary surfactant-associated protein
A precursor"

Region 1..15 /gene="SFTPA1" /region_name="Signal" /note="POTENTIAL."

Region 12 /gene="SFTPA1" /region_name="Variant" /note="S -> P."

Region 16..247 /gene="SFTPA1" /region_name="Mature chain"

/note="PULMONARY SURFACTANT-ASSOCIATED PROTEIN A."

Region 27..99 /gene="SFTPA1" /region_name="Domain" /note="COLLAGEN-LIKE."

Region 57..60 /gene="SFTPA1" /region_name="Conflict" /note="GPMG -> APWA
(IN REF. 2)."

Region 152..247 /gene="SFTPA1" /region_name="Domain" /note="C-TYPE LECTIN
(SHORT FORM)."

Bond bond(154,245) /gene="SFTPA1" /bond_type="disulfide" /note="BY
SIMILARITY."

Site 206 /gene="SFTPA1" /site_type="glycosylation" /note="N-LINKED (GLCNAC...)
(PROBABLE)."

Bond bond(223,237) /gene="SFTPA1" /bond_type="disulfide" /note="BY
SIMILARITY."

ORIGIN 1 mllslaltl isapasdtcd tkdvcigspg ipgtpgshgl pgrdgrdgvk gdpggppgpmg
61 ppggmpglpg rdgligapgv pgergdkgep gergppglpa yldeelqatl helrhhalqs
121 igvlslqgsm kavgekfst ngqsvnfai revcaraggr iavprseen eaiasivker
181 ntyaylglae gptagdfyyl dgdpvnytnw ypgeprgqgr ekcvemytdg.kwndknclqy
241 rlvicef

SEQ ID NO: 84

NP_033186 surfactant associated protein D [Mus musculus]
gi|6677921|ref|NP_033186.1|[6677921]

sig_peptide 1..19

mat_peptide 20..374 /product="surfactant associated protein D"

Region 260..373 /region_name="C-type lectin (CTL) or carbohydrate-recognition
domain (CRD)" /note="CLECT" /db_xref="CDD:smart00034"

Region 271..374 /region_name="Lectin C-type domain" /note="lectin_c"
/db_xref="CDD:pfam00059"

CDS 1..374 /gene="Sftpd" /coded_by="NM_009160.1:43..1167"

/db_xref="LocusID:20390" /db_xref="MGD:109515"

ORIGIN 1 mlplfmlvl lvqplnlgla emkslsqrsv pntctlvmscs ptenglprgd grdgregprg
61 ekgdpglpdp mglsglqgpt gpvgpkge sagepgpkge rglsppglp gipgpagkeg

121 psgkqgnigp qgkpgpkgea gpkgevgapg mqqstgakgs tgpkgergap
 gvqgapgnag
 181 aagpagpagp qgapgsrgpp glkgdrvgpg drgikgesgl pdsaalrqm ealkgklqrl
 241 evafshyqka alfpdgrsvg dkifrtadse kpfedaqemc kqaggqlasp rsatenaaiq
 5 301 qlitahnkaa flsmtdvgte gkftyptgep lvsynwapge pnnnggaenc veiftngqwn
 361 dkacgeqrly icef

SEQ ID NO: 85

1B08C Chain C, Lung Surfactant Protein D (Sp-D) (Fragment)

10 gi|6573321|pdb|1B08|C[6573321]

FEATURES Location/Qualifiers source 1..158 /organism="Homo sapiens"
 /db_xref="taxon:9606"

15 SecStr 13..36 /sec_str_type="helix" /note="helix 7"
 Region 38..158 /region_name="Domain 3" /note="NCBI Domains"
 SecStr 39..44 /sec_str_type="sheet" /note="strand 21"
 SecStr 45..51 /sec_str_type="sheet" /note="strand 22"
 SecStr 53..56 /sec_str_type="sheet" /note="strand 23"
 20 SecStr 57..67 /sec_str_type="helix" /note="helix 8"
 Bond bond(64,156) /bond_type="disulfide"
 SecStr 77..90 /sec_str_type="helix" /note="helix 9"
 SecStr 93..96 /sec_str_type="sheet" /note="strand 24"
 Het join(bond(100),bond(100),bond(100),bond(104),bond(104),
 bond(104),bond(127),bond(132),bond(133)) /heterogen="(CA, 8)"
 25 Het join(bond(104),bond(133),bond(133),bond(133)) /heterogen="(CA, 9)"
 SecStr 107..110 /sec_str_type="sheet" /note="strand 25"
 SecStr 112..115 /sec_str_type="sheet" /note="strand 26"
 Het join(bond(124),bond(126),bond(132),bond(144),bond(145),
 bond(145),bond(145),bond(145),bond(145),bond(145),
 30 bond(145),bond(145),bond(145),bond(145),bond(145),
 bond(145),bond(145),bond(145),bond(145),bond(145),
 bond(145),bond(145),bond(145),bond(145),bond(145), bond(145)) /heterogen="(CA, 7)"
 SecStr 133..139 /sec_str_type="sheet" /note="strand 27"
 35 Bond bond(134,148) /bond_type="disulfide"
 SecStr 141..147 /sec_str_type="sheet" /note="strand 28"
 SecStr 150..158 /sec_str_type="sheet" /note="strand 29"
 ORIGIN 1 eaeagsvasl rqqvealqgg vqhlqaafsq ykkvelfpng qsvgekifkt agfvkpftea
 61 qltctqaggq lasprsaen aalqqlvvak neaafismtd sktegkftyp tgeslvysnw
 40 121 apgepnddgg sedcveiftn gkwndracge krlvvcf

SEQ ID NO: 86

1B08B Chain B, Lung Surfactant Protein D (Sp-D) (Fragment)

45 gi|6573320|pdb|1B08|B[6573320]

FEATURES Location/Qualifiers source 1..158 /organism="Homo sapiens"
 /db_xref="taxon:9606"

SecStr 11..34 /sec_str_type="helix" /note="helix 4"
 Region 37..158 /region_name="Domain 2" /note="NCBI Domains"
 50 SecStr 39..44 /sec_str_type="sheet" /note="strand 11"
 SecStr 45..51 /sec_str_type="sheet" /note="strand 12"
 SecStr 53..56 /sec_str_type="sheet" /note="strand 13"
 SecStr 57..67 /sec_str_type="helix" /note="helix 5"

Bond bond(64,156) /bond_type="disulfide"
 SecStr 77..90 /sec_str_type="helix" /note="helix 6"
 SecStr 93..96 /sec_str_type="sheet" /note="strand 14"
 SecStr 97..100 /sec_str_type="sheet" /note="strand 15"
 5 Het join(bond(100),bond(100),bond(100),bond(104),bond(104),
 bond(104),bond(127),bond(132),bond(133)) /heterogen="(CA, 5)"
 Het join(bond(104),bond(133),bond(133),bond(133)) /heterogen="(CA, 6)"
 SecStr 107..110 /sec_str_type="sheet" /note="strand 16"
 Het join(bond(124),bond(126),bond(132),bond(144),bond(145), bond(145))
 10 /heterogen="(CA, 4)"
 SecStr 133..139 /sec_str_type="sheet" /note="strand 17"
 Bond bond(134,148) /bond_type="disulfide"
 SecStr 141..147 /sec_str_type="sheet" /note="strand 18"
 SecStr 150..153 /sec_str_type="sheet" /note="strand 19"
 15 SecStr 154..158 /sec_str_type="sheet" /note="strand 20"
 ORIGIN 1 eaeagsvasl rqqvealqgg vqhlqaafsq ykkvelfpng qsvgekifkt agfvkpfta
 61 qltctqaggq lasprsaaen aalqlvvak neaafismtd sktegkftyp tgeslvysnw
 121 apgepnddgg sedcveiftn gkwndracge krlvvcef

 20 SEQ ID NO: 87
 1B08A Chain A, Lung Surfactant Protein D (Sp-D) (Fragment)
 gi|6573319|pdb|1B08|A[6573319]

 FEATURES Location/Qualifiers source 1..158 /organism="Homo sapiens"
 25 /db_xref="taxon:9606"
 SecStr 10..36 /sec_str_type="helix" /note="helix 1"
 Region 38..158 /region_name="Domain 1" /note="NCBI Domains"
 SecStr 39..44 /sec_str_type="sheet" /note="strand 1"
 SecStr 45..51 /sec_str_type="sheet" /note="strand 2"
 30 SecStr 53..56 /sec_str_type="sheet" /note="strand 3"
 SecStr 57..67 /sec_str_type="helix" /note="helix 2"
 Bond bond(64,156) /bond_type="disulfide"
 SecStr 77..90 /sec_str_type="helix" /note="helix 3"
 SecStr 93..96 /sec_str_type="sheet" /note="strand 4"
 35 SecStr 97..100 /sec_str_type="sheet" /note="strand 5"
 Het join(bond(100),bond(100),bond(100),bond(104),bond(104),
 bond(104),bond(127),bond(132),bond(133)) /heterogen="(CA, 2)"
 Het join(bond(104),bond(133),bond(133),bond(133)) /heterogen="(CA, 3)"
 SecStr 107..110 /sec_str_type="sheet" /note="strand 6"
 40 Het join(bond(124),bond(126),bond(132),bond(144),bond(145), bond(145))
 /heterogen="(CA, 1)"
 SecStr 133..139 /sec_str_type="sheet" /note="strand 7"
 Bond bond(134,148) /bond_type="disulfide"
 SecStr 141..147 /sec_str_type="sheet" /note="strand 8"
 45 SecStr 150..153 /sec_str_type="sheet" /note="strand 9"
 SecStr 154..158 /sec_str_type="sheet" /note="strand 10"
 ORIGIN 1 eaeagsvasl rqqvealqgg vqhlqaafsq ykkvelfpng qsvgekifkt agfvkpfta
 61 qltctqaggq lasprsaaen aalqlvvak neaafismtd sktegkftyp tgeslvysnw
 121 apgepnddgg sedcveiftn gkwndracge krlvvcef
 50
 SEQ ID NO: 88
 NP_060049 deleted in malignant brain tumors 1 isoform c precursor [Homo sapiens]
 gi|8923740|ref|NP_060049.1|[8923740]

sig_peptide 1..25
 mat_peptide 26..2403 /product="deleted in malignant brain tumors 1 isoform c"
 Region 102..202 /region_name="Scavenger receptor Cys-rich" /note="SR"
 5 /db_xref="CDD:SR"
 Region 105..202 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 234..334 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 10 Region 237..334 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 363..463 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 15 Region 366..463 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 484..584 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 487..584 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 20 Region 594..692 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 595..692 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 25 Region 723..823 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 726..823 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 852..952 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 30 Region 855..952 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 983..1083 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 35 Region 986..1083 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 1112..1212 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 1115..1212 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 40 Region 1241..1341 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 1244..1341 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 45 Region 1370..1470 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 1373..1470 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 1499..1599 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 50 Region 1502..1599 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 1630..1730 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"

Region 1633..1730 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 1756..1867 /region_name="Domain first found in C1r, C1s, uEGF, and bone
 morphogenetic protein." /note="CUB" /db_xref="CDD:CUB"
 5 Region 1756..1864 /region_name="CUB domain" /note="CUB"
 /db_xref="CDD:pfam00431"
 Region 1873..1976 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 1885..1976 /region_name="Scavenger receptor cysteine-rich domain"
 10 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 1998..2106 /region_name="Domain first found in C1r, C1s, uEGF, and bone
 morphogenetic protein." /note="CUB" /db_xref="CDD:CUB"
 Region 1998..2104 /region_name="CUB domain" /note="CUB"
 /db_xref="CDD:pfam00431"
 15 Region 2117..2371 /region_name="Zona pellucida-like domain"
 /note="zona_pellucida" /db_xref="CDD:pfam00100"
 Region 2117..2368 /region_name="Zona pellucida (ZP) domain" /note="ZP"
 /db_xref="CDD:ZP"
 20 CDS 1..2403 /gene="DMBT1" /coded_by="NM_017579.1:107..7318" /note="isoform
 c is encoded by transcript variant 3" /db_xref="LocusID:1755"
 /db_xref="MIM:601969"
 ORIGIN 1 mgistvilem clwggqvist ggwiprttdy aslipsevpl dttvaegspf pseltlesta
 25 61 aegspisles tlestvaegs lipsestles tvaegsdsgl alrlvngdgr cqgrveilyr
 121 gswgtvcdds wdtnanvvc rqlgcgwams apgnawfggg sgpiaddvr csghesylws
 181 cphngwlshn cghhedagvi csaaqpqstl rpeswpvris ppvptegses slalrlvngg
 241 drcgrvevl yrgswgtvcd dywdtnanv vcrqlgcgwa msapgnafgq qsggpivldd
 301 vrcsghesyl wscphngwlt hncghsedag vicsaplsrp tpsdtwpts hastagpess
 361 lalrlvnggd rcqgrvevly rgswgtvcdd swdtsdanvv crqlgcgwat sapgnarfqq
 421 gsgpivlddv rcsghesylw scphngwlsh ncqhsedagv icsdtlptit lpastvgsses
 30 481 slalrlvngg drcqgrvevl yrgswgtvcd dswdtnanv vcrqlgcgwa mlapgnarfq
 541 qsggpivldd vrcsgnesyl wscphngwls hncghsedag vicsgpessl alglvnggdr
 601 cqgrvevlyr gswgtvcdds wdtnanvvc rqlgcgwats apgnarfqqg sgpiaddvr
 661 csghesylws cpnngwlshn cghhedagvi csaaqsrstp rpdltititl ppstvgsses
 721 ltlrlvngsd rcqgrvevly rgswgtvcdd swdtnanvv crqlgcgwat sapgnarfqq
 35 781 gsgpivlddv rcsghesylw scphngwlsh ncghhedagv icsvsqsrt pspdtwptsh
 841 astagsessl alrlvnggdr cqgrvevlyr gswgtvcdds wdtsdanvvc rllgcgwats
 901 apgnarfqqg sgpiaddvr csgyesylws cphngwlshn cqhsedagvi csahswstp
 961 spdtlptitl pastvgsses lalrlvnggd rcqgrvevly qgswgtvcdd swdtnanvv
 1021 crqlgcgwam sapgnarfqq gsgpivlddv rcsghesylw scphngwlsh ncghsedagv
 40 1081 icsasqsrt pspdtwptsh astagsessl alrlvnggdr cqgrvevlyr gswgtvcddy
 1141 wdtnanvvc rqlgcgwams apgnarfqqg sgpiaddvr csghesylws cphdgwlshn
 1201 cghhedagvi csasqsqptp spdtwptsha stagsessla lrlvnggdrc qgrvevlyrg
 1261 pwgtvcddyw dtnanvvcr qlgcgwatsa pgnarfqqgs gpivlddvrc sghesylwsc
 1321 phngwlshnc ghhedagvic sasqsqptps pdtwptshas tagsesslal rlvnggdrcq
 45 1381 grvevlyrgs wgtvcddywd tnanvvcrq lgcgwatsap gsarfgqgsg pialddvrcs
 1441 ghesylwscp hngwlshncg hhedagvics asqsqptpsp dtwptsrast agsestlral
 1501 lvnggdrcrg rvevlyqgsw gtvccdywdt ndanvvcrql gcgwamsapg naqfgqgsgp
 1561 ivlddvrcsg hesylwscph ngwlshncgh hedagvicsa aqsqstprpd twlttnlpal
 1621 tvgsesslal rlvnggdrcr grvevlyrgs wgtvcddswd tnanvvcrq lgcgwamsap
 50 1681 gnarfqqgs pivlddvrcs gnesylwscp hkgwlthncg hhedagvics atqinstttd
 1741 wwhtpttita rpssncggfl fyasgtfssp syayypnna kcvweievns gyrlngfsn
 1801 lkleahhncs fdyveifdgs insslllgki cndtrqifts synrmtihfr sdisfntgf
 1861 lawynsfpsd alrlvnlms syglcagrve iyhggwtgtv cddswtiqea evvcrqlgcg

5

10

1921 ravsalignay fgsgsgpiti dvecsgtes tlwqcmrgw fshncnhred agvicsgnhl
 1981 stpaplinit.rpntdyscgg flsqpsgdfs spfypgnypn nakcvwdiev qnnyrvtvif
 2041 rdvqleggc n ydyievdgp yrsspliav cdgargsfts ssnfmsirfi sdhsitrgrf
 2101 raeyyspsn dstnllclpn hmqaasvsry lqslgfsasd lvistwngyy ecrpqitpnl
 2161 viftipysgc gtfkqadndt idysnfltaa vsggiikrrt dlrihvscrm lqntwvdtmy
 2221 iandtihvan ntiqveevqy gnfdvnisyf tsssflypvt srpyyvdlnq dlyvqaeilh
 2281 sdavltlfvd.tcvaspysnd ftsltydlir sgcvrddtyg pysspslrira rfrfrafhfl
 2341 nrfpsvylrc kmvvcraydp ssrccrgcvl rskrdvgsyq ekvdvvlgpi qlqtpprree
 2401 epr

15

20

25

30

35

40

45

50

SEQ ID NO: 89

NP_015568 deleted in malignant brain tumors 1 isoform b precursor [Homo sapiens]
 gi|6633801|ref|NP_015568.1||6633801]

sig_peptide 1..25

mat_peptide 26..2413 /product="deleted in malignant brain tumors 1. isoform b"

Region 102..202 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"

Region 105..202 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"

Region 234..334 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"

Region 237..334 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"

Region 363..463 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"

Region 366..463 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"

Region 494..594 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"

Region 497..594 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"

Region 602..702 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"

Region 605..702 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"

Region 733..833 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"

Region 736..833 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"

Region 862..962 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"

Region 865..962 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"

Region 993..1093 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"

Region 996..1093 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"

Region 1122..1222 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"

Region 1125..1222 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"

Region 1251..1351 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 1254..1351 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 5 Region 1380..1480 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 1383..1480 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 10 Region 1509..1609 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 1512..1609 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 1640..1740 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 15 Region 1643..1740 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 1766..1877 /region_name="Domain first found in C1r, C1s, uEGF, and bone
 morphogenetic protein." /note="CUB" /db_xref="CDD:CUB"
 20 Region 1766..1874 /region_name="CUB domain" /note="CUB"
 /db_xref="CDD:pfam00431"
 Region 1883..1986 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 1895..1986 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 25 Region 2008..2116 /region_name="Domain first found in C1r, C1s, uEGF, and bone
 morphogenetic protein." /note="CUB" /db_xref="CDD:CUB"
 Region 2008..2114 /region_name="CUB domain" /note="CUB"
 /db_xref="CDD:pfam00431"
 30 Region 2127..2381 /region_name="Zona pellucida-like domain"
 /note="zona_pellucida" /db_xref="CDD:pfam00100"
 Region 2127..2378 /region_name="Zona pellucida (ZP) domain" /note="ZP"
 /db_xref="CDD:ZP"
 CDS 1..2413 /gene="DMBT1" /coded_by="NM_007329.1:107..7348"
 /db_xref="LocusID:1755" /db_xref="MIM:601969"
 35 ORIGIN 1 mgistvilem clwgqvlt ggwiprttdy aslipsevpl dqtvaegspf psestlesta
 61 aegspisles tlestvaegs lipsestles tvaegsdsgl alrlvngdgr cqgrveilyr
 121 gswgtvcdds wdtnanvvc rqlgcgwams apgnawfgqg sgpiaddvr csghesylws
 181 cphngwlshn cghgedagvi csaaqpqstl rpeswpvris ppvptegses slalrlvngg
 241 drcgrvèvl yrgswgtvcd dywdtnanv vcrqlgcgwa msapgnafgq qsgpividd
 40 301 vrcsghesyl wscphngwlt hncghsedag vicsapqsrp tpspdtwpts hastagpess
 361 lalrlvnggd rcqgrveily rsgwtvcdd swdtdanvv crqlgcgwat sapgnarfqq
 421 gsgpividdv rcsghesylw scphngwlsh ncghsedagv icsaahswst pspdtlptit
 481 lpastvgsses slalrlvngg drcqgrvèvl yrgswgtvcd dswdtnanv vcrqlgcgwa
 541 mlapgnarfqq qsgpividd vrcsgnesyl wscphngwls hncghsedag vicsgpessl
 45 601 alrlvngdgr cqgrveilyr gswgtvcdds wdtnanvvc rqlgcgwams apgnarfqqg
 661 sgpiaddvr csghesylws cpnngwlshn cghhedagvi csaaqsrstp rptltititl
 721 ppstvgsses lrlrlvngsd rcqgrveily rsgwtvcdd swdtdanvv crqlgcgwam
 781 sapgnarfqq gsgpividdv rcsghesylw scphngwlsh ncghhedagv icsvsqsrt
 841 pspdtwptsh astagsessl alrlvngdgr cqgrveilyr gswgtvcdds wdtdanvvc
 50 901 rqlgcgwats apgnarfqqg sgpiaddvr csghesylws cphngwlshn cghsedagvi
 961 csaahswstp spdtlptitl pastvgsses lalrlvnggd rcqgrveily qsgwtvcdd
 1021 swdtdanvv crqpgcgwam sapgnarfqq gsgpividdv rcsghesypw
 scphngwlsh

1081 ncghsedagv icsasqsrt pspdtwptsh astagsessl alrlvnggdr cqgrvevlyr
 1141 gswgtvcddy wdtnanvvc rqlcgwams apgnarfqqg sgpiivddvr csghesylws
 1201 cphngwlshn cghhedagvi csasqsqtp spdtwptsha stagsessla lrlvnggdr
 1261 qgrvevlyrg swgtvcddyw dtndanvvc rqlcgwatsa pgnarfqqgs gpivddvr
 5 1321 sghesylwsc phngwlshnc ghhedagvic sasqsqtps pdtwptshas tagsessla
 1381 rlvnggdrq grvevlyrgs wgtvcddywd tndanvvc rqlcgwatsa pgnarfqqgs
 1441 pivddvrscs ghesylwscph hngwlshncg hhedagvics asqsqtpsp dtwptsrast
 1501 agsestlrl lrvnggdrq rvevlyqgs wgtvcddywd tndanvvc rqlcgwatsa pgn
 1561 naqfgqgsq gpivddvrscs ghesylwscph ngwlshncg hhedagvicsa aqsqstprp
 10 1621 twlttnlpal tvgsessla rlvnggdrq grvevlyrgs wgtvcddswd tndanvvc r
 1681 lcgwatsa pgnarfqqgs pivddvrscs ghesylwscph hkgwlthncg hhedagvics
 1741 atqinsttd wwhptttta rpsncggf fyasgtfssp sypayypna kcvweievns
 1801 gyrinlgfsn kcleahhncs fdyveifdgs lnslllgki cndtrqfts synrmtihfr
 1861 sdisfntgf lawynsfpsd atrlvnlms syglcagrve iyhggtwgtv cddswtiqea
 15 1921 evvcrqlcg ravsalgnay fgsgsgpiti dvecsgtes tlwqcrnrgw fshncnhred
 1981 agvicsgnhl stpafinit rptdyscgg flsqpsgdfs spfypgnypn nakcvwdiev
 2041 qnnrvvtvif rdvqleggc ydyievfdg yrsspliarv cdgargstfs snfmsirfi
 2101 sdhsitrrg raeyyspsn dstnlclpn hmqasvsrsy lqslgfsasd lvistwnggy
 2161 ecrpqtprl viftipysgc gtfkqadndt idysnflta vsggiikrt dlrihvscm
 20 2221 lqntwvdtmy landtihvan ntiqueevqy gnfdvnisfy tsssflypvt srpyvdlng
 2281 dlyvqaeilh sdavltifvd tcvapsysnd ftslydlir sgcvrddtyg pysspslria
 2341 rfrfrahfl nrpsvylrc kmvvcraydp ssrvcrgcvl rskrdvgsyq ekdvdlvgpi
 2401 qlqtpprree epr

25 SEQ ID NO: 90
 NP_004397 deleted in malignant brain tumors 1 isoform a precursor [Homo sapiens]
 gi|4758170|ref|NP_004397.1|[4758170]

sig_peptide 1..25
 30 mat_peptide 26..1785 /product="deleted in malignant brain tumors 1 isoform a"
 Region 102..202 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 105..202 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 35 Region 234..334 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 237..334 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 363..463 /region_name="Scavenger receptor Cys-rich" /note="SR"
 40 /db_xref="CDD:SR"
 Region 366..463 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 494..594 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 45 Region 497..594 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 623..723 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 626..723 /region_name="Scavenger receptor cysteine-rich domain"
 50 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 752..852 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"

48.

Region 755..852 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 881..981 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 5 Region 884..981 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 1012..1112 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 10 Region 1015..1112 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 Region 1138..1249 /region_name="Domain first found in C1r, C1s, uEGF, and bone
 morphogenetic protein." /note="CUB" /db_xref="CDD:CUB"
 Region 1138..1246 /region_name="CUB domain" /note="CUB"
 /db_xref="CDD:pfam00431"
 15 Region 1255..1358 /region_name="Scavenger receptor Cys-rich" /note="SR"
 /db_xref="CDD:SR"
 Region 1267..1358 /region_name="Scavenger receptor cysteine-rich domain"
 /note="SRCR" /db_xref="CDD:pfam00530"
 20 Region 1380..1488 /region_name="Domain first found in C1r, C1s, uEGF, and bone
 morphogenetic protein." /note="CUB" /db_xref="CDD:CUB"
 Region 1380..1486 /region_name="CUB domain" /note="CUB"
 /db_xref="CDD:pfam00431"
 Region 1499..1751 /region_name="Zona pellucida-like domain"
 /note="zona_pellucida" /db_xref="CDD:pfam00100"
 25 Region 1499..1750 /region_name="Zona pellucida (ZP) domain" /note="ZP"
 /db_xref="CDD:ZP"
 CDS 1..1785 /gene="DMBT1" /coded_by="NM_004406.1:107..5464"
 /db_xref="LocusID:1755" /db_xref="MIM:601969"
 ORIGIN 1 mgistvilem clwggqvlt ggwiprttdy aslipsevpl dqtvaegspf psestlesta
 30 61 aegspisles tlestvaegs lipsestles tvaegsdsl alrlvngdgr cqgrveilyr
 121 gswgtvcdds wdtnanvvc rqlgcgwams apgnawfgqg sgpiatddvr csghesylws
 181 cphngwlshn cghgedagvi csaaqpqstl rpeswpvris ppvptegses slalrlvngg
 241 drcgrvevl yrgswgtvcd dywdtnanv vcrqlgcgwa msapgnafg qsgspividd
 301 vrcsghesyl wscphngwlt hncghsedag vicsapqsrp tpspdtwpts hastagpess
 35 361 lalrlvnggd rcqgrveily rgswwgtvcd swdtsdanvv crqlgcgwat sapgnarfqq
 421 gsgpividdv rcsygesylw scphngwlsh ncqhsedagv icsaahswst pspdtlptit
 481 lpastvgsses slalrlvngg drcqgrvevl yqgswwgtvcd dswdtnanv vcrqpgcgwa
 541 msapgnarfqq qsgspividd vrcsghesyp wscphngwls hncghsedag vicsasqsrp
 601 tpspdtwpts hastagsses lalrlvnggd rcqgrveily rgswwgtvcd ywdtnanv
 40 661 crqlgcgwam sapgnarfqq gsgpividdv rcsghesylw scphngwlsh ncghhedagv
 721 icsasqsqpt pspdtwptsh astagssesl alrlvnggdrc cqgrveilyr gswgtvcddy
 781 wdtnanvvc rqlgcgwats apgnarfqqg sgpiatddvr csghesylws cphngwlshn
 841 cghhedagvi csasqsqptp spdtwptstra stagsestla lrlvnggdrc rgveilyqq
 901 swgtvcddyw dtndanvvc rqlgcgwamsa pgnafgqgs gpividdvrc sghesylwsc
 45 961 phngwlshnc ghhedagvic saasqsqstr pdtwittnlp altvgssesl alrlvnggdrc
 1021 crgrveilyr gswgtvcdds wdtnanvvc rqlgcgwams apgnarfqqg sgpiatddvr
 1081 csgnesylws cphkgwlshn cghhedagvi csatqinstt tdwwhptttt tarpssncgg
 1141 ffyasgtfs spsyayypn nakcvweiev nsyrlngf snkleahhn csfdyveifd
 1201 glnsslllg kicndtrqif tssynrmtih frsdisfnt glawynsf sdatlrlvnl
 50 1261 nssyglcagr veiyhggtwg tvcdswtiq eaevvcrqlg cgravsaln ayfgsgsgpi
 1321 tldvecsgt estlwqcmr gwfnhcnhr edagvicsgn hlstpapfln itrptdysc
 1381 ggflsqpsgd fsspfypgny pnnakcvwdi evqnnyrvtv ifrdvqlegg cnydyieifd
 1441 gpyrssplia rvcdgargsf tsssnfmsir fisdhsitr gfraeyysp sndstnlcl

1501 pnhmqasvsr sylqslgfsa sdvlistwng yyecrpqitp nlviftipys gcgtfkqadn
 1561 dtidysnflit aavsggiikr rtdlrihvsc rmlqntwvdt myiandtiHV anntiqveev
 1621 qygnfdvnis fytsssflyp vtsrpyyvdI nqdlyvqaei lHsdavltlf vdtcvaspys
 1681 ndftsltydl irsgcvrddt ygpysspslr iarfrfrah flnrfpsvyl rckmvvcray
 5 1741 dpssrcyrgc vlrskrdvgs yqekvdvlg piqlqtprr eeepR

SEQ ID NO: 91

LNBOC1 pulmonary surfactant protein C – bovine

gi|7428752|pir||LNBOC1[7428752]

10 FEATURES Location/Qualifiers source 1..34 /organism="Bos taurus"
 /db_xref="taxon:9913"

Protein 1..34 /product="pulmonary surfactant protein C" /note="pulmonary surfactant protein PSP-6"

Site 4 /site_type="binding" /note="palmitate (Cys) (covalent)"

15 Site 5 /site_type="binding" /note="palmitate (Cys) (covalent)"

ORIGIN 1 lipccpvnik rllivvvvv lllvvivgal lmgI

SEQ ID NO: 92

LNDGC1 pulmonary surfactant protein C – dog gi|7428750|pir||LNDGC1[7428750]

20 FEATURES Location/Qualifiers source 1..35 /organism="Canis familiaris"
 /db_xref="taxon:9615"

Protein 1..35 /product="pulmonary surfactant protein C"

Site 5 /site_type="binding" /note="palmitate (Cys) (covalent)"

ORIGIN 1 lgpcfpssI krllivvvi vlvvvivga llmgI //

SEQ ID NO: 93

JN0450 conglutinin precursor – bovine gi|346501|pir||JN0450[346501]

FEATURES Location/Qualifiers source 1..371 /organism="Bos taurus"

30 /db_xref="taxon:9913"

Protein 1..371 /product="conglutinin precursor" /note="C3b-binding protein"

Region 1..20 /region_name="domain" /note="signal sequence"

Region 21..371 /region_name="product" /note="conglutinin"

Region 46..214 /region_name="region" /note="collagen-like"

35 Site 63 /site_type="binding" /note="carbohydrate (Lys) (covalent)"

Site 63 /site_type="modified" /note="5-hydroxylysine (Lys)"

Region 75..371 /region_name="product" /note="conglutinin-N"

Site 78 /site_type="modified" /note="4-hydroxyproline (Pro)"

Site 87 /site_type="binding" /note="carbohydrate (Lys) (covalent)"

40 Site 87 /site_type="modified" /note="5-hydroxylysine (Lys)"

Site 96 /site_type="modified" /note="4-hydroxyproline (Pro)"

Site 99 /site_type="binding" /note="carbohydrate (Lys) (covalent)"

Site 99 /site_type="modified" /note="5-hydroxylysine (Lys)"

Site 108 /site_type="modified" /note="4-hydroxyproline (Pro)"

45 Site 111 /site_type="modified" /note="4-hydroxyproline (Pro)"

Site 129 /site_type="modified" /note="4-hydroxyproline (Pro)"

Site 132 /site_type="modified" /note="4-hydroxyproline (Pro)"

Site 135 /site_type="binding" /note="carbohydrate (Lys) (covalent)"

Site 135 /site_type="modified" /note="5-hydroxylysine (Lys)"

50 Site 141 /site_type="binding" /note="carbohydrate (Lys) (covalent)"

Site 141 /site_type="modified" /note="5-hydroxylysine (Lys)"

Site 147 /site_type="modified" /note="4-hydroxyproline (Pro)"

Site 153 /site_type="modified" /note="4-hydroxyproline (Pro)"

Site 159 /site_type="binding" /note="carbohydrate (Lys) (covalent)"
 Site 159 /site_type="modified" /note="5-hydroxylysine (Lys)"
 Site 162 /site_type="binding" /note="carbohydrate (Lys) (covalent)"
 Site 162 /site_type="modified" /note="5-hydroxylysine (Lys)"
 5 Site 171 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 195 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 198 /site_type="binding" /note="carbohydrate (Lys) (covalent)"
 Site 198 /site_type="modified" /note="5-hydroxylysine (Lys)"
 Site 210 /site_type="binding" /note="carbohydrate (Lys) (covalent)"
 10 Site 210 /site_type="modified" /note="5-hydroxylysine (Lys)"
 Region 248..369 /region_name="domain" /note="C-type lectin homology #label
 LCH"
 Site 337 /site_type="binding" /note="carbohydrate (Asn) (covalent)"
 ORIGIN 1 millplsvll lltqpwrsig aemttsqki lanactlvmc splesglpgh dgqdgrecph
 15 61 gekgdpqspg pagragrpgw vgpigpkgdn gfvgepgpkg dtgprgppgm
 pgpagregps
 121 gkqgsmgppg tpgpkgetgp kggvgapgiq gfpgpsglkg ekgapgetga pgragvtgps
 181 gaigpqgpgs argppglkgd rgdpgetgak gesglavna lkqrvtildg hlrrfqnafs
 241 qykkavlfpd qgavgekifk tagavksysd aeqlcreakg qlasprssae neavtqmvr
 20 301 qeknaylsmn distegrfty pteilvyn wadgpnnsd egqpencvei fpdgkwndvp
 361 cskqllvice f

SEQ ID NO: 94
 A45225 pulmonary surfactant protein D precursor – human
 25 gi|346375|pir||A45225[346375]
 FEATURES Location/Qualifiers source 1..375 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 Protein 1..375 /product="pulmonary surfactant protein D precursor" /note="SP-D"
 Region 1..20 /region_name="domain" /note="signal sequence"
 30 Region 21..375 /region_name="product" /note="pulmonary surfactant protein D"
 Region 21..45 /region_name="domain" /note="non-collagenous"
 Region 46..222 /region_name="domain" /note="collagenous"
 Site 90 /site_type="binding" /note="carbohydrate (Asn) (covalent)"
 Region 223..375 /region_name="domain" /note="non-collagenous"
 35 Region 254..373 /region_name="domain" /note="C-type lectin homology #label
 LCH"
 Bond bond(281,373) /bond_type="disulfide"
 Bond bond(351,365) /bond_type="disulfide"
 ORIGIN 1 mlflfalslv lltqplgyle aemktyshrt mpsactlvmc ssvesglpgr dgrdgregpr
 40 61 gekgdpplpg aagqagmpgq agpvgpkgdn gsvgepgpkg dtgpsgppgp
 pgvpgpagre
 121 galgkqgnig pqgkpgpkge agpkgevgap gmqgsagarg lagpkgergv
 pgergvpgnt
 181 gaagsagamg pqgspgargp pglkgdkgip gdkgakgesg lpdvaslrqq vealqgqvqh
 45 241 lqaafsqqyk velfpngqsv gekifktagf vkpftaqll ctqaggqlas prsaaenaal
 301 qqlvvaknea aflsmtskt egkftyptge slvysnwapg epnddggsed cveiftngkw
 361 ndraccgekrl vvcef

SEQ ID NO: 95
 50 LNHUC pulmonary surfactant protein C precursor, long splice form – human
 gi|71983|pir||LNHUC[71983]

FEATURES Location/Qualifiers source 1..197 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 Protein 1..197 /product="pulmonary surfactant protein C precursor, long splice form"
 /note="3.7 kDa:surfactant polypeptide; pulmonary surfactant protein SP5; pulmonary
 5 surfactant proteolipid SP-C; pulmonary surfactant proteolipid SPL(pVal)"
 Region 1..197 /region_name="product" /note="pulmonary surfactant protein C
 precursor, short splice form"
 Region 1..145 /region_name="product" /note="pulmonary surfactant protein C
 precursor, short splice form"
 10 Region 1..23 /region_name="domain" /note="propeptide"
 Region 24..58 /region_name="product" /note="pulmonary surfactant protein C"
 Site 28 /site_type="binding" /note="palmitate (Cys) (covalent)"
 Site 29 /site_type="binding" /note="palmitate (Cys) (covalent)"
 Region 152..197 /region_name="product" /note="pulmonary surfactant protein C
 15 precursor, short splice form" ORIGIN 1 mdvgskevln esppdysaap rgrfgipccp vhlkrllivv
 vvvllivvvi vgallmglhm
 61 sqkhtemvle msigapeaaqqlrlsehlvt tatfsigstg lvvydyqqll iaykpapgtc
 121 cyimkiapes ipslealnrlk vhnfqmecsll qakpavptsk lgqaegrdrag sapsggdpa
 181 lgmavntlcg evplyyi
 20
 SEQ ID NO: 96
 LNDGPS pulmonary surfactant protein A precursor – dog
 gi|71970|pir||LNDGPS[71970]
 FEATURES Location/Qualifiers source 1..248 /organism="Canis familiaris"
 25 /db_xref="taxon:9615"
 Protein 1..248 /product="pulmonary surfactant protein A precursor"
 /note="pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated
 protein PSP-A"
 Region 1..17 /region_name="domain" /note="signal sequence"
 30 Region 18..248 /region_name="product" /note="pulmonary surfactant protein A"
 Site 20 /site_type="binding" /note="carbohydrate (Asn) (covalent)"
 Region 28..102 /region_name="region" /note="collagen-like"
 Site 30 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Region 127..246 /region_name="domain" /note="C-type lectin homology #label
 35 LCH"
 Site 207 /site_type="binding" /note="carbohydrate (Asn) (covalent)"
 ORIGIN 1 mwlrlclal tllmvsgien ntkdvcvgnp gipgtpgshg lpgrdgrdgv kgdpgppgpl
 61 gppggmpgghp gpngmtgapg vagergekge pgergppglp asldeelqtl lhdrlrhqilq
 121 tmgvslshes llvgrkvfs sgaqsinfnd iqelcagagg qiaapmspee neavasivkk
 40 181 yntyaylgiv espdsqdfqy mdgapvnytn wypgeprgrg keqcvmeytd qgwnknclq
 241 yrlaicef
 SEQ ID NO: 97
 LNHUPS pulmonary surfactant protein A precursor (genomic clone) – human
 45 gi|71967|pir||LNHUPS[71967]
 FEATURES Location/Qualifiers source 1..248 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 Protein 1..248 /product="pulmonary surfactant protein A precursor (genomic clone)"
 /note="alveolar proteinosis protein; pulmonary surfactant 32K apoprotein; pulmonary
 50 surfactant-associated protein (PSP-A)"
 Region 1..20 /region_name="domain" /note="signal sequence"
 Region 21..248 /region_name="product" /note="pulmonary surfactant protein A"
 Bond bond(26) /bond_type="disulfide" /note="interchain"

Region 28..100 /region_name="domain" /note="collagenous"
 Site 30 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 33 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 36 /site_type="modified" /note="4-hydroxyproline (Pro)"
 5 Site 42 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 51 /site_type="modified" /note="5-hydroxylysine (Lys)"
 Site 57 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 63 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 76 /site_type="modified" /note="4-hydroxyproline (Pro)"
 10 Site 79 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 82 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 88 /site_type="modified" /note="5-hydroxylysine (Lys)"
 Site 91 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Site 97 /site_type="modified" /note="4-hydroxyproline (Pro)"
 15 Region 127..246 /region_name="domain" /note="C-type lectin homology #label LCH"
 Bond bond(155,246) /bond_type="disulfide"
 Site 207 /site_type="binding" /note="carbohydrate (Asn) (covalent)"
 Bond bond(224,238) /bond_type="disulfide"
 20 ORIGIN 1 mwlclplaln lmaasgavc evkdvcvgs gipgtpgshg lpgrhgrdgl kgdlgpppgm
 61 gppgempcpp gndglpgag ipgecgckge pgergppglp ahldelqat lhdfrhqlq
 121 trgalisqgs imtvgekvfs sngqsitfda iqeacaragg riavprnpee neaiafvkk
 181 yntyayvglt egpspgdfry sdgtpvnytn wyrgepagrg keqcvemytd gqwndmcl
 241 srlticef
 25
 SEQ ID NO: 98
 A53570 collectin-43 – bovine gi|1083017|pir||A53570[1083017]
 30
 FEATURES Location/Qualifiers source 1..301 /organism="Bos taurus"
 /db_xref="taxon:9913"
 Protein 1..301 /product="collectin-43" /note="lectin CL-43"
 Region 177..299 /region_name="domain" /note="C-type lectin homology #label LCH"
 35
 ORIGIN 1 eemdvyskt ltdpctlvvc appadsrlgh dgrdgkegpq gekgdpgppg mpgpagregp
 61 sgrqgsmgpp gtpgpkgepg peggvgapgm pgsppaglk gergapppg
 aigpqgpsga
 121 mgppglkgdr gdpgekgarg etsvlevdtl rqrnrllege vqrlqnvitq yrkavlfpdg
 181 qavgekifkt agavksysda eqlcreakgq lasprssaen eavtqlvrak nkhaylsmd
 40 241 iskegkftyp tggsl dysnw apgepnrak degpenclei ysdgnwndie creerlvce 301
 f
 SEQ ID NO: 99
 S33603 surfactant protein D – bovine gi|423283|pir||S33603[423283]
 45
 FEATURES Location/Qualifiers source 1..369 /organism="Bos taurus"
 /db_xref="taxon:9913"
 Protein 1..369 /product="surfactant protein D"
 Region 248..367 /region_name="domain" /note="C-type lectin homology #label LCH"
 50
 ORIGIN 1 mlllplsvll lltqpwrsig aemkiysqkt manactivmc sppedglpgr dgrdgregpr
 61 gekgdpgspg pagragmpgp agpiglkgdn gsagepgpkg dtgppgppgm
 pgpagregps

121 gkqgsmgppg tpgpkgtgp kggvgagpiq gspgpaglk ergapgdpga
 pgragapgr
 181 gaigpqgpg argppglkgd rgtppgergak gesglaevna lrqrvgileg qlrlqnafs
 241 qykkamlfpn grsvgekifk tvgsektfqd aqciqtqagg qlpsprsgae nealtqlata
 5 301 qnkaafisms dtrkegtfiy ptgeplvysn wapqepnndg gsencveifp ngkwndkvcg
 361 eqrivicef

SEQ ID NO: 100
 AAF28384 lung surfactant protein A [Sus scrofa]
 10 gi|6782434|gb|AAF28384.1|AF133668_1[6782434]
 FEATURES Location/Qualifiers source 1..116 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 Protein <1..116 /product="lung surfactant protein A" /function="involved in the innate
 immune system and lipid homeostasis within the lung" /name="collectin; SPA; SP-A"
 15 CDS 1..116 /gene="SFTPA" /coded_by="AF133668.1:<1..353"
 ORIGIN 1 avgekvfstn gqsvafdvir elcaraggri aaprspeene aiasivkkhn tyaylgiveg
 61 ptagdffyl d gtpvnytnwy pgeprgrgke kcvernytdgq wdrncqqyr laicef

SEQ ID NO: 101
 20 AAF22145 lung surfactant protein D precursor; SPD; SP-D; CP4 [Sus scrofa]
 gi|6760482|gb|AAF22145.2|AF132496_1[6760482]
 sig_peptide 1..20
 mat_peptide 21..378 /product="lung surfactant protein D"
 CDS 1..378 /gene="SFTPD" /coded_by="AF132496.2:44..1180"
 25 ORIGIN 1 mlllplsvli lltqpprslg aemktysqra vanacalvmc spmenglpgr dgrdgregpr
 61 gekgdpglpg avgragmpgl agpvpgkgn gstgepgakg digpcgppgp
 pgipgpagke
 121 gpsgqqgnig ppgtppgkge tgpkgvgal gmqgstgarg paglkgerga pgergapgsa
 181 gaagpagatg pqgpgargp pglkgdrgpp gergakgesg lpgitalrqv vetlqqqvqr
 30 241 lqkafsqykk velfpnrgv gekifktggf ektfqdaqqv ctqaggqmas prseteneal
 301 sqlvtaqnka aflsmt dikt egnftyptge plvyanwapg epnnnggssg aencveifpn
 361 gkwndkacge lrlvicef

SEQ ID NO: 102

P15783 PULMONARY SURFACTANT-ASSOCIATED PROTEIN C (SP-C)
(PULMONARY SURFACTANT-ASSOCIATED PROTEOLIPID SPL(VAL))
gi|131422|sp|P15783|PSPC_BOVIN[131422]

5 FEATURES Location/Qualifiers source 1..34 /organism="Bos taurus"
/db_xref="taxon:9913"
gene 1..34 /gene="SFTPC" /note="SFTP2"
Protein 1..34 /gene="SFTPC" /product="PULMONARY SURFACTANT-
10 ASSOCIATED PROTEIN C"
Site 4 /gene="SFTPC" /site_type="lipid-binding" /note="PALMITATE (BY
SIMILARITY)."
Site 5 /gene="SFTPC" /site_type="lipid-binding" /note="PALMITATE (BY
SIMILARITY)."
15 Region 21 /gene="SFTPC" /region_name="Conflict" /note="L -> V (IN REF. 2)."
Region 26 /gene="SFTPC" /region_name="Conflict" /note="I -> V (IN REF. 2)."
Region 28..34 /gene="SFTPC" /region_name="Conflict" /note="GALLMGL ->
IGAMLAM (IN REF. 2)."
ORIGIN 1 lipccpvnik rllivvvvv llvvvivgal lmgI

20 SEQ ID NO: 103

P35246 PULMONARY SURFACTANT-ASSOCIATED PROTEIN D PRECURSOR
(SP-D) (PSP-D)
gi|464485|sp|P35246|PSPD_BOVIN[464485]

25 FEATURES Location/Qualifiers source 1..369 /organism="Bos taurus"
/db_xref="taxon:9913"
gene 1..369 /gene="SFTPD" /note="SFTP4"
Protein 1..369 /gene="SFTPD" /product="PULMONARY SURFACTANT-
30 ASSOCIATED PROTEIN D PRECURSOR"
Region 1..20 /gene="SFTPD" /region_name="Signal" /note="BY SIMILARITY."
Region 21..369 /gene="SFTPD" /region_name="Mature chain"
/note="PULMONARY SURFACTANT-ASSOCIATED PROTEIN D."
Region 46..216 /gene="SFTPD" /region_name="Domain" /note="COLLAGEN-LIKE."
35 Site 78 /gene="SFTPD" /site_type="hydroxylation" /note="(BY SIMILARITY)."
Site 87 /gene="SFTPD" /site_type="hydroxylation" /note="(BY SIMILARITY)."
Site 90 /gene="SFTPD" /site_type="glycosylation" /note="POTENTIAL."
Site 96 /gene="SFTPD" /site_type="hydroxylation" /note="(BY SIMILARITY)."
Site 99 /gene="SFTPD" /site_type="hydroxylation" /note="(BY SIMILARITY)."
40 Site 165 /gene="SFTPD" /site_type="hydroxylation" /note="(BY SIMILARITY)."
Site 171 /gene="SFTPD" /site_type="hydroxylation" /note="(BY SIMILARITY)."
Region 217..248 /gene="SFTPD" /region_name="Domain" /note="COILED COIL
(POTENTIAL)."
Region 273..369 /gene="SFTPD" /region_name="Domain" /note="C-TYPE LECTIN
45 (SHORT FORM)."
Bond bond(275,367) /gene="SFTPD" /bond_type="disulfide" /note="BY
SIMILARITY."
Bond bond(345,359) /gene="SFTPD" /bond_type="disulfide" /note="BY
SIMILARITY."
50 ORIGIN 1 mlllplsvll lltqpwrslg aemkiysqkt manactlvmc sppedglpgr dgrdgregpr
61 gekgdpgspg pagragmpgp agpiglkgn gsagepgpkg dtgppppgm
pgpagregps

121 gkqgsmgppg tpgpkgtgp kggvgapgiq gspgpaglk ergapgepga
 pgragapgpa
 181 gaigpqgpg argppglkgd rgtppgergak gesglaevna lrqrvgileg qlrlqnafs
 241 qykkamlfpn grsvgekifk tvgsektfqd aqciqtqagg qlpsprsgae nealtqlata
 5 301 qnkaafilms dtrkegtfiy ptgeplvysn wapqepnndg gsencveifp ngkwndkvcg
 361 eqlvicef

SEQ ID NO: 104

P42916 COLLECTIN-43 (CL-43) gi|1168967|sp|P42916|CL43_BOVIN[1168967]

10 FEATURES Location/Qualifiers source 1..301 /organism="Bos taurus"
 /db_xref="taxon:9913"

Protein 1..301 /product="COLLECTIN-43"

Region 29..142 /region_name="Domain" /note="COLLAGEN-LIKE (G-X-Y)."

15 Region 202..301 /region_name="Domain" /note="C-TYPE LECTIN (SHORT
 FORM)."

Bond bond(204,299) /bond_type="disulfide" /note="BY SIMILARITY."

Bond bond(277,291) /bond_type="disulfide" /note="BY SIMILARITY."

ORIGIN 1 eemdvysekt ltdpctlvvc appadslrgh dgrdgkegpq gekgdpqpgp mppgpagregp

61 sgrqgsmgpp gtpgpkgepg peggvgapgm pgsppaglk gergapggg

20 aigpqgpgsa

121 mgppglkgdr gdpgekgarg etsvlevdtl rqrnrlege vqrlqivtq yrkavifpdg

181 qavgekifkt agavksysda eqlcreakgq lasprssaen eavtqlvrak nkhaylsmnd

241 iskegkftyp tggsl dysnw apgepgnrak degpenclei ysdgnwndie creerlvce

301 f

25 SEQ ID NO: 105

CAB56155 DMBT1/8kb.2 protein [Homo sapiens]

gi|5912464|emb|CAB56155.1|[5912464]

sig_peptide 1..26

30 mat_peptide 26..2412 /product="DMBT1/8kb.2 protein"

CDS 1..2412 /gene="DMBT1" /coded_by="AJ243212.1:107..7345"

/note="Sequence is an alternative splice form of the DMBT1 gene that is expressed
 in human adult trachea. Isoforms of DMBT1 are identical to the collectin binding
 protein gp-340. Full-length cDNA clone contains 1 bp deletions in codons 100 and

35 1751, that were corrected by comparison with the genomic exons"

ORIGIN 1 mgistvilem clwgvqlst ggwiprttdy aslipsevpl dttvaegspf pseltlestv

61 aegspisles tlettvaegs lipsestles tvaegsdsgl alrlvngdgr cqgrveilyr

121 gswgavcdds wdtdanvvc rqlgcgwams apgnawfgqg sgpiatddvr csghesylws

181 cphngwlsn cghgedagvi csaaqpqstl rpeswpvris ppvptegses slalrvngg

40 241 drcrgrvevl yrgswgtvcd dywdtdanv vcrqlgcgwa msapgnaqfg qgsgpivldd

301 vrcsghesyl wscphngwlt hncghsedag vicsapqsrp tpspdtwpts hastagpess

361 lalrvnggd rcqgrvevly rgswgtvcdd swdtdanvv crqlgcgwat sapgnarfqq

421 gsgpivlddv rcsghesylw scphngwlsn ncqhsedagv icsaahswst pspdtlptit

481 lpastvgsses slalrvngg drcqgrvevl yrgswgtvcd dswdtdanv vcrqlgcgwa

45 541 mlapgnarfqq qgsgpivldd vrcsgnesyl wscphngwls hncghsedag vicsgpessl

601 alrlvnggdrc qqgrvevlyr gswgtvcdds wdtdanvvc rqlgcgwams apgnarfqqg

661 sgpiatddvr csghesylws cpnngwlsn cghhedagvi csaaqsrstp rpdltitit

721 ppstvgsses ltlrvngsd rcqgrvevly rgswgtvcdd swdtdanvv crqlgcgwat

781 sapgnarfqq gsgpivlddv rcsghesylw scphngwlsn ncqhsedagv icsvsqsrt

50 841 pspdtwptsh astagpessl alrlvnggdrc qqgrvevlyr gswgtvcdds wdtdanvvc

901 rqlgcgwats apgnarfqqg sgpiatddvr csgyesylws cphngwlsn cghsedagvi

961 csaaahswstp spdtlptitl pastvgsses lalrvnggd rcqgrvevly qgswgtvcdd

1021 swdtdanvv crqlgcgwam sapgnarfqq gsgpivldd rcsghesylw scphngwlsn

1081 ncghsedagv icsasqsrt pspdtwptsh astagsessl alrlvnggdr cqgrvevlyr
 1141 gswgtvcddy wdtnanvac rqlgcgwams apgnarfqq sgpiivddvr csghesylws
 1201 cphngwlshn cghhedagvi csasqsqtp spdtwptsha stagsessla lrlvnggdr
 1261 qgrvevlyrg swgtvcddyw dtndanvvr qlgcgwatsa pgnarfqqgsg gpividdvr
 5 1321 sghesylwsc phngwlshnc ghhedagvic sasqsqtps pdtwptshas tagssessla
 1381 rlvggdrcc grvevlyrgs wgtvcddywd tndanvvrq lgcgwatsap gnarfqqsg
 1441 pividdvrcs ghesylwscph hngwlshncg hhedagvics afqsqtpsp dtwptsrast
 1501 agsestlaln lvggdrccrg rvevlyqgs wgtvcddywd tndanvvrq lgcgwamsap
 1561 naqfgqsgp ividdvrcsg hepylwscph ngwlshncg hddagvicsa aqsqstprpd
 10 1621 twltlnpal tvgsessla rlvggdrcc grvevlyrgs wgtvcddswd tndanvvrq
 1681 lgcgwamsap gnarfqqsg pivlvdvrcs ghesylwscph hkgwlthncg hhedagvics
 1741 atqinstttd wwhptttta rpsnccgfi fyasgtfss syayypnna kcvweievns
 1801 gyrlngfsn kkleahhncs fdyveifdgs lnslllgki cndtrqfts synrmtihfr
 1861 sdisfntgf lawynsfpsd alrlvnlms syglcagrv ihggwtgav cddswtqea
 15 1921 evvcrqlgcg ravsalignay fgsggpitl dvecsgtes tlwqcrnrgw fshncnhred
 1981 agvicsgnhl stpaplnt rpnyscgfi lqpsgdfss pfypgnypnn akcvwdievq
 2041 nnyrvtivr dvqleggcny dyievfdgy rsspliarvc dgargstss snfmsirfis
 2101 dhsitrrgr aeyyspsnd stnlclpnh mqasvsrsyl qlsgfsasdl vistwngyye
 2161 crpqtprlv iftipysgcy tkqadndti dynlltaav sggiikrrtd lrlhvscrm
 20 2221 qntwvdtmyi andtihvann tiqveevqy nfdvnisfy sssflypvt rpyvdlndq
 2281 lyvqaeilhs davltilvdt cvaspsndf tslydlirs gcvrddtygp ysspslriar
 2341 frfrahfln rpsvylrck mvvcraydps srcyrgcvlr skrdvgsyqe kvdvlglpiq
 2401 lqtprrree pr

 25 SEQ ID NO: 106
 AAD49696 gp-340 variant protein [Homo sapiens]
 gi|5733598|gb|AAD49696.1|AF159456_1[5733598]
 FEATURES Location/Qualifiers source 1..2413 /organism="Homo sapiens"
 /db_xref="taxon:9606" /chromosome="10" /map="10q25.3-26.1"
 30 Protein 1..2413 /product="gp-340 variant protein" /name="scavenger receptor
 cysteine-rich protein SRCR" /note="putative receptor for SP-D"
 CDS 1..2413 /gene="DMBT1" /coded_by="AF159456.1:107..7348"
 ORIGIN 1 mgistvilem clwgqvist ggwiprttdy aslipsevpl dqtvaegspf psestlesta
 61 aegspisles tlestvaege lipsestles tvaegsdsgl alrlvngdgr cqgrveilyr
 35 121 gswgtvcdds wdtnanvvc rqlgcgwams apgnawfgq sgpiavddvr csghesylws
 181 cphngwlshn cghhedagvi csaaqpqstl rpeswpvris ppvptegses slalrlvngg
 241 drcgrvevi yrgswgtvcd dywdtnanv vcrqlgcgwa msapgnafqg qsgpiavdd
 301 vrcsghesyl wscphngwlt hncghsedag vicsapqsrp tpsdwtwpts hastagpess
 361 lalrlvnggd rcqgrvevly rgswwgtvcd swdtsdanv crqlgcgwat sapgnarfqq
 40 421 gsgpiavddv rcsghesylw scphngwlsh ncghsedagv icsaahswst pspdtlptit
 481 lpastvgsses slalrlvngg drcqgrvevi yrgswgtvcd dswdtnanv vcrqlgcgwa
 541 mlapgnarfq qsgpiavdd vrcsgnesyl wscphngwls hncghsedag vicsgpessl
 601 alrlvnggdr cqgrvevlyr gswgtvcdds wdtnanvvc rqlgcgwams apgnarfqqg
 661 sgpiavddvr csghesylws cpnngwlshn cghhedagvi csaaqsrst rpdltititl
 45 721 ppstvgsses lrlvngsd rcqgrvevly rgswwgtvcd swdtnanv crqlgcgwam
 781 sapgnarfqq gsgpiavddv rcsghesylw scphngwlsh ncghhedagv icsvsqsrt
 841 pspdtwptsh astagsessl alrlvnggdr cqgrvevlyr gswgtvcdds wdtsdanvvc
 901 rqlgcgwats apgnarfqq sgpiavddvr csgyesylws cphngwlshn cghsedagvi
 961 csaaahswst pspdtlptitl pastvgsses lalrlvnggd rcqgrvevly qgswwgtvcd
 50 1021 swdtnanv crqpgcgwam sapgnarfqq gsgpiavddv rcsghesypw scphngwlsh
 1081 ncghsedagv icsasqsrt pspdtwptsh astagsessl alrlvnggdr cqgrvevlyr
 1141 gswgtvcddy wdtnanvvc rqlgcgwams apgnarfqq sgpiavddvr csghesylws
 1201 cphngwlshn cghhedagvi csasqsqtp spdtwptsha stagsessla lrlvnggdr

1261 qgrvevlyrg swgtvcddyw dtndanvvcrl qlgcgwatsa pgnarfqqgs gpivddvrc
 1321 sghesylwsc phngwishnc ghhedagvic sasqsqtps pdtwptshas tagsesslal
 1381 rlvnggdrcq grvevlyrgs wgtvcddywd tndanvvcrl lcggwatsap gnarfqqgs
 1441 pivddvrcs ghesylwscph hngwishncg hhedagvics asqsqtpsp dtwptsrast
 5 1501 agsèstlalr lvnggdrcrg rvevlyqgs wgtvcddywdt ndanvvcrl lcggwamsap
 1561 naqfgqgsqp ivddvrcsg hesylwscph ngwishncgh hedagvicsa aqsqstprpd
 1621 twlttnlpal tvgsesslal rlvnggdrcr grvevlyrgs wgtvcddswd tndanvvcrl
 1681 lcggwamsap gnarfqqgs pivddvrcs ghesylwscph hngwishncg hhedagvics
 1741 atqinstttd wwhptttta rpssnccgfl fyasgtfssp syayypnna kcvweievns
 10 1801 gyrinlgfsl klahhncs fdyveifdgs lnslllgi cndtrqifts synrmtihfr
 1861 sdisfntqtf lawynsfpsd atlrlnlns syglcagrve iyhggtwgtv cddswtiqea
 1921 evvcrqlgcg ravsalignay fgsgsgpiti ddvecsgtes tlwqcrnrgw fshncnhred
 1981 agvicsgnhl stpaplinit rpntdyscgg flsqpsgdfs spfypgnypn nakcvwdiev
 2041 qnnrvtvif rdvqleggc ydyievfdgp yrsspliarv cdgargsfts ssnfmsirfi
 15 2101 sdhsitrgrf raeyyspsn dstnlclpn hmqaavrsy lqslgfsasd livistwnggy
 2161 ecrpqtlnl viftipysgc gtfkqadndt idysnflta vsggiikrrt dlrihvscrm
 2221 lqntwvdtmy iandtihvan ntiqveevqy gnfdvnisfy tsssflypvt srpyyvdlnq
 2281 dlyvqaeilh sdavltfvd tcvaspysnd ftstydilr sgcvrddtyg pysspslria
 2341 rfrfrahfil nrfpsvylrc kmvvcraydp ssrccrgcvl rskrdvgsyq ekvdivlgi
 20 2401 qlqtpprree.epr

SEQ ID NO:107

AAD31380 surfactant protein D precursor [Mus musculus]

gi|4877556|gb|AAD31380.1|AF047742_1[4877556]

sig_peptide 1..19

mat_peptide 20..374 /product="surfactant protein D"

CDS 1..374 /gene="Sftp4"

/coded_by="join(AF047741.1:5705..5900,AF047742.1:312..428,

AF047742.1:669..785,AF047742.1:1112..1228,

30 AF047742.1:1977..2093,AF047742.1:3162..3245, AF047742.1:5010..5386)"

ORIGIN 1 mlplfmlvl lvqplgnlga emkslsqrsv pntctlvms ptenglpggd grdgregprg

61 ekgdpglpgp mglsglqgpt gvpqpkgeag sagepgpkge rglsgppglp gipgpagkeg

121 psgkqgnigp qgkpgpkgea gpkgevgapg mqgstgakgs tgpkgergap

gvqgapgnag

35 181 aagpagpagp qgapgsrgpp glkgdrvgp drgikgesgl pdsaalrqm ealkgklql

241 evafshyqka alfpdgrsvg dkifrtadse kpfedaqemc kqaggqlasp rsatenaaiq

301 qlitahnkaa flsmtdivgte gkftyptgep lvysnwapge pnnnggaenc veiftngqwn

361 dkacgeqlv icef

40 SEQ ID NO: 108

B61249 pulmonary surfactant protein C – dog gi|539712|pir||B61249[539712]

FEATURES Location/Qualifiers source 1..35 /organism="Canis familiaris"

/db_xref="taxon:9615"

Protein 1..35 /product="pulmonary surfactant protein C"

45 ORIGIN 1 lgipcfpsl krllivvvi vlvvvviva llmgl

SEQ ID NO: 109

S00609 pulmonary surfactant protein C – bovine gi|89749|pir||S00609[89749]

50 FEATURES Location/Qualifiers source 1..34 /organism="Bos taurus"

/db_xref="taxon:9913"

Protein 1..34 /product="pulmonary surfactant protein C" /note="pulmonary surfactant protein PSP-6"

58

Site 4 /site_type="binding" /note="palmitate (Cys) (covalent)"

Site 5 /site_type="binding" /note="palmitate (Cys) (covalent)"

ORIGIN 1 lipccpvnik rllivvvvv llvvvivgal lmgf

5 SEQ ID NO: 110

A43628 pulmonary surfactant protein A - human (fragments)

gi|280854|pir||A43628[280854]

FEATURES Location/Qualifiers source 1..35 /organism="Homo sapiens"

/db_xref="taxon:9606"

10 Protein 1..35 /product="pulmonary surfactant protein A"

ORIGIN 1 gqsitfdagk eqcvemytdg qwndrnclyl ticef

SEQ ID NO: 111

AAB48076 Surfactant protein B (SP-B) [Oryctolagus cuniculus]

15 gi|1850933|gb|AAB48076.1|[1850933]

FEATURES Location/Qualifiers source 1..370 /organism="Oryctolagus cuniculus"

/db_xref="taxon:9986" /tissue_type="liver"

Protein 1..370 /product="Surfactant protein B (SP-B)"

CDS 1..370 /gene="SP-B"

20 /coded_by="join(U40853.1:2194..2263,U40853.1:2591..2718,

U40853.1:2941..3012,U40853.1:3257..3382,

U40853.1:3590..3727,U40853.1:3925..4014,

U40853.1:6043..6226,U40853.1:6421..6581,

25 U40853.1:7266..7346,U40853.1:7829..7891)" /note="Surfactant protein B (SP-B) is

a key component of lung surfactant, a surface active material secreted by type II

epithelial cells of lung alveolus; SP-B maintains biophysical properties and

physiological function of surfactant; Pulmonary surfactant associated protein"

ORIGIN 1 makshlppwl llllptlclg pgtavwatasp lacaqgpew cqsleqalqc kalghclqev

30 61 wghvgaddlc qecqdivnil tkmtkeafq dtirkflehe cdvlpkliv pqchhldvy

121 fplityfls qinakaicqh lgcqpqspe pldplpdkl vptllgalp akpgphtqdl

181 saqrfpiplp lcwlcrllk riqamipkgv lamavaqvch vvplvvggic qclaerytvi

241 llevlghvl pqlvcglvr cssvdsigqv pptlealpgc wlpqdpecpl cmsvttqarn

301 iseqtrpqav yhadlssqld kqecqfvel htpqlsls rgwdaraicq algacvatls

361 plqciqspfh

35

SEQ ID NO: 112

1901176A surfactant protein A gi|382753|prf||1901176A[382753]

FEATURES Location/Qualifiers source 1..247 /organism="Oryctolagus cuniculus"

/db_xref="taxon:9986"

40 ORIGIN 1 mllslatl isapasdtdc tkdvcigspg ipgtpgshgl pgrdgrdgvk gdpग्प्ग्प्ग्

61 ppgmpgplpg rdgligapgv pgergdkgep gergppglpa yldeelqatl helrhhalqs

121 igvlslqgsm kavgekfst ngqsvnfai revcaraggr iaivkevprs leeneaiasr

181 ntyaylglae gptagdfyyl dgdpvnytnw ypgeprggr ekcvemytdg kwndknclqy

241 rviccf

45

SEQ ID NO: 113

CAA53510 lung surfactant protein D [Bos taurus]

gi|415939|emb|CAA53510.1|[415939]

sig_peptide 1..20

50 mat_peptide 21..369 /product="lung surfactant protein D"

CDS 1..369 /coded_by="X75911.1:102..121" /db_xref="SWISS-PROT:P35246"

ORIGIN 1 mlllplsvll lltqpwrslg aemkiysqkt manactlymc sppedglpgr dgrdgregpr

61 gekgdpqspg pagragmpgp agpiglkgn gsagepgpkg dtgppgppgm
 pgpagregps
 121 gkqgsmgppg tpgpkgtgp kggvgapgiq gspgpaglk ergapgepga
 pgragapga
 5 181 gaigpqgpg argppglkgd rgtppgergak gesglaevna lqrvlgileg qlqlqnafs
 241 qykkamlfpn grsvgekifk tvgsektfqd aqqictagg qlpsprsgae nealtqlata
 301 qnkaafisms dtrkegtfiy ptgeplvysn wapqepnndg gsencveifp ngkwndkvvcg
 361 eqlvicef

10 SEQ ID NO: 114
 CAA53511 collectin-43 [Bos taurus] gi|499385|emb|CAA53511.1|[499385]

15 FEATURES Location/Qualifiers source 1..301 /organism="Bos taurus"
 /db_xref="taxon:9913" /tissue_type="liver" /clone_lib="lambda gt 11"
 Protein 1..301 /product="collectin-43"
 mat_peptide 1..301 /product="collectin-43"
 CDS 1..301 /coded_by="X75912.1:<1..906" /db_xref="SWISS-PROT:P42916"
 ORIGIN 1 eemdvyxekt ltdpctivvc appadslrgh dgrdgkegpq gekgdpqppg mpgpagregp
 20 61 sgrqgsmgpp gtpgpkgepg pegvgapgm pgspgpaglk gergapggg
 aigpqgpgsa
 121 mgppglkgdr gdpgekgarg etsvlevdtl rqrmmlege vqlqnivtq yrkavlfpdg
 181 qavgekifkt agavksysda eqlcreakgq lasprssaen eavtqlvrak nkhaylsmdnd
 241 iskegkftyp tggslidysnw apgepgnrak degpenclei ysdgnwndie creerlvce
 301 f

25 SEQ ID NO: 115
 CAA46152 lung surfactant protein D [Homo sapiens]
 gi|34767|emb|CAA46152.1|[34767]

30 sig_peptide 1..20
 mat_peptide 21..375 /product="lung surfactant protein D"
 CDS 1..375 /gene="hsp-D" /coded_by="X65018.1:172..1299" /db_xref="SWISS-
 PROT:P35247"
 ORIGIN 1 mlflfalslv lltqplgyle aemktyshrt tpsactlvmc ssvesglpgr dgrdgregpr
 35 61 gekgdpqlpg aagqagmpgq agpvpgkgn gsvgepgpkg dtgpgppgp
 pgvppgagre
 121 gplgkqgnig pqgkpgpkge agpkgevgap gmqgsagarg lagpkgergv
 pgergvpgna
 181 gaagsagamg pqgspgargp pglkgdkgip gdkgakgesg lpdvaslrqq vealqgqvqh
 40 241 lqaafsqqyk velfpngqsv gekifktagf vkpfteaql ctqaggqlas prsaaenaal
 301 qqlvvaknea aflsmtskt egkftptge slvysnwapg epnddggsed cveifngkw
 361 ndraccgkrl vvcef

45 SEQ ID NO: 116
 AAA92788 lung surfactant protein C [Rattus norvegicus]
 gi|595282|gb|AAA92788.1|[595282]
 FEATURES Location/Qualifiers source 1..194 /organism="Rattus norvegicus"
 /db_xref="taxon:10116" /clone="sp-c" /tissue_type="liver"
 Protein 1..194 /product="lung surfactant protein C"
 50 CDS 1..194 /gene="sp-c"
 /coded_by="join(U07796.1:1673..1714,U07796.1:2841..2999,
 U07796.1:3252..3377,U07796.1:3598..3707,U07796.1:4053..4200)"
 ORIGIN 1 mdmgskevlm esppdystgp rsqfripccp vhlkrlliv vvvlvvvvi vgallmgllhm

60

61 sqkhtemvle msiggapetq krlalsehtd tiatfsigst givlydyqrl ltaykpapgt
 121 ycyimkmape sipslealar kfkfnqakss tptsklgqee ghsagsdsds sgrdlafigl
 181 avstlcgvlp lyi

5 SEQ ID NO: 117

AAA31468 surfactant protein A [*Oryctolagus cuniculus*]
 gi|431446|gb|AAA31468.1|[431446]

10 FEATURES Location/Qualifiers source 1..247 /organism="Oryctolagus cuniculus"
 /strain="New Zealand White" /db_xref="taxon:9986" /tissue_type="liver"
 /dev_stage="adult"

Protein 1..247 /product="surfactant protein A"

CDS 1..247 /coded_by="join(L19387.1:3864..4032,L19387.1:4241..4360,
 L19387.1:5010..5087,L19387.1:5533..5909)"

15 ORIGIN 1 mllslaltl isapasdtd tkdvcigspg ipgtpgshgl pgrdgrdvk gdpgrpapwa
 61 ppgmpglpg rdgligapgv pgergdkgep gergppglpa yldeelqatl helrhhalqs
 121 igvlslqgsm kavgekifst ngqsvnfai revcaraggr iavprleen eaiaivker
 181 ntyaylglae gptagdfyyl dgdpvnytnw ypgeprgqgr ekcvemytdg kwndknclqy
 241 rlvicef

20

Mannose binding lectin

25 SEQ ID NO: 1

NP_034897 mannan-binding lectin serine protease 2 [*Mus musculus*]
 gi|6754642|ref|NP_034897.1|[6754642]

sig_peptide 1..15

30 mat_peptide 16..185 /product="mannan-binding lectin serine protease 2"
 Region 28..137 /region_name="Domain first found in C1r, C1s, uEGF, and bone
 morphogenetic protein" /note="CUB" /db_xref="CDD:smart00042"
 Region 28..134 /region_name="CUB domain" /note="CUB"
 /db_xref="CDD:pfam00431"

35 Region 138..180 /region_name="Calcium-binding EGF-like domain"
 /note="EGF_CA" /db_xref="CDD:smart00179"

variation 172 /allele="I" /allele="V" /db_xref="dbSNP:3167338"

CDS 1..185 /gene="Masp2" /coded_by="NM_010767.1:32..589"

/db_xref="LocusID:17175" /db_xref="MGD:1330832"

40 ORIGIN 1 mrlilfigll wslvatllgs kwpepvfgrl vspgfpekya dhqdrswtlt appgyrlrl
 61 fthfdlelsy rceydfvklsgtkvlatlc gqestdteqa pgndtfyslg pslkvtfhds
 121 ysnekpftgf eafyaaedyd ecrvslgdsd pcdhychnyl ggyycscrag yvlhqnhkhtc
 181 seqsl

45 SEQ ID NO: 2

AAH10760 Similar to mannanose binding lectin, serum (C) [*Mus musculus*]
 gi|14789670|gb|AAH10760.1|[14789670]

50 source 1..244 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090"
 /clone="MGC:18500 IMAGE:4212216" /tissue_type="Liver, normal. 5 month old
 male mouse." /clone_lib="NCI_CGAP_Li9" /lab_host="DH10B" /note="Vector:
 pCMV-SPORT6"

Protein 1..244 /product="Similar to mannanose binding lectin, serum (C)"

61

CDS 1..244 /coded_by="BC010760.1:192..926"

ORIGIN 1 msiftsfill cvvtvyaet ltegvqncsp vvtcsspgln gfpkgdgrdg akgekgepgg
 61 glrglqgppg kvgtpppgn pglkgavgpk gdrgrdraefd tseidseiaa lrselralrn
 121 wvlfslsekv gkkyfvssvk kmsldrvkal csefggsvat prnaeensa qkvakdiayl
 181 gitdvrvegs fedltgnrvr ytnwndgepn ntgdgedcvv ilgngkwndv pcsdsflaic
 241 efsd

SEQ ID NO: 3

AAH21762 mannose binding lectin, liver (A) [Mus musculus]

gi|18256010|gb|AAH21762.1|[18256010]

source 1..239 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090"
 /clone="MGC:30242 IMAGE:5132514" /tissue_type="Liver, normal. 5 month old
 male mouse." /clone_lib="NCI_CGAP_Li9" /lab_host="DH10B" /note="Vector:
 pCMV-SPORT6"

Protein 1..239 /product="mannose binding lectin, liver (A)"

/db_xref="LocusID:17194"

CDS 1..239 /coded_by="BC021762.1:75..794" /db_xref="LocusID:17194"

ORIGIN 1 millpllpvl lcvtvsyssg sqtcedtlkt csviacgrdg rdgpkgekge pgqglrglqg
 61 ppgklgppgs vsgpspgpk gqkgdhgdnr aieeklanme aeirilkskl qltnklhafs
 121 mgkksgkklf vtnhekmfks kvkslctelq gtvaiprae enkaiqevat giaflgitde
 181 ategqfmyvt ggrltylnwk kdepnnhsg edcvliidng lwndiscqas fkavcefp

SEQ ID NO: 4

Q9NPY3 Complement component C1q receptor precursor (Complement component
 1, q subcomponent, receptor 1) (C1qRp) (C1qR(p)) (C1q/MBL/SPA receptor) (CD93
 antigen) (CDw93) gi|21759074|sp|Q9NPY3|CD93_HUMAN[21759074]

source 1..652 /organism="Homo sapiens" /db_xref="taxon:9606

gene 1..652 /gene="C1QR1" /note="CD93"

Protein 1..652 /gene="C1QR1" /product="Complement component C1q receptor precursor"

Region 1..21 /gene="C1QR1" /region_name="Signal"

Region 22..652 /gene="C1QR1" /region_name="Mature chain"

/note="COMPLEMENT COMPONENT C1Q RECEPTOR.

Region 22 /gene="C1QR1" /region_name="Conflict" /note="T -> V (IN AA SEQUENCE)."

Region 24..580 /gene="C1QR1" /region_name="Domain" /note="EXTRACELLULAR (POTENTIAL)."

Region 32..174 /gene="C1QR1" /region_name="Domain" /note="C-TYPE LECTIN."

Region 36 /gene="C1QR1" /region_name="Conflict" /note="C -> T (IN AA SEQUENCE)."

Region 38..39 /gene="C1QR1" /region_name="Conflict" /note="TA -> RI (IN AA SEQUENCE)."

Region 155 /gene="C1QR1" /region_name="Conflict" /note="S -> N (IN REF. 1)."

Region 186 /gene="C1QR1" /region_name="Conflict" /note="G -> A (IN AA SEQUENCE)."

Region 260..301 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 1."

Bond bond(264,275) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILARITY."

Bond bond(271,285) /gene="C1QR1" /bond_type="disulfide" /note="BY SIMILARITY."

Bond bond(287,300) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Region 302..344 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 2."
 Bond bond(306,317) /gene="C1QR1" /bond_type="disulfide" /note="BY
 5 SIMILARITY."
 Bond bond(311,328) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Region 318 /gene="C1QR1" /region_name="Variant" /note="V -> A.
 /FTId=VAR_013573."
 10 Site 325 /gene="C1QR1" /site_type="glycosylation" /note="N-LINKED (GLCNAC...)
 (POTENTIAL)."
 Bond bond(330,343) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 15 Region 345..384 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 3,
 CALCIUM-BINDING (POTENTIAL)."
 Bond bond(349,358) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Bond bond(354,367) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 20 Bond bond(369,383) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Region 385..426 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 4,
 CALCIUM-BINDING (POTENTIAL)."
 Bond bond(389,400) /gene="C1QR1" /bond_type="disulfide" /note="BY
 25 SIMILARITY."
 Bond bond(396,409) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Bond bond(411,425) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 30 Region 427..468 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 5,
 CALCIUM-BINDING (POTENTIAL)."
 Bond bond(431,443) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Bond bond(439,452) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 35 Bond bond(454,467) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Region 492 /gene="C1QR1" /region_name="Conflict" /note="S -> A (IN AA
 SEQUENCE)."
 Region 496 /gene="C1QR1" /region_name="Conflict" /note="R -> Q (IN AA
 40 SEQUENCE)."
 Region 504 /gene="C1QR1" /region_name="Conflict" /note="R -> G (IN AA
 SEQUENCE)."
 Region 541 /gene="C1QR1" /region_name="Conflict" /note="P -> S (IN REF. 1)."
 Region 581..601 /gene="C1QR1" /region_name="Transmembrane region"
 45 /note="POTENTIAL."
 Region 594..601 /gene="C1QR1" /region_name="Domain" /note="POLY-LEU."
 Region 602..652 /gene="C1QR1" /region_name="Domain" /note="CYTOPLASMIC
 (POTENTIAL)."
 ORIGIN 1 matsmgllll lllltqpga gtgadteavv cvgtacytah sgksaaeeq nhonqnggnl
 50 61 atvkskeeeq hvqrilaql rreaaltarm skfwiglqre kgkclpdlsp lkgfswvggg
 121 edtpysnwhk elrnsciskr cvslldlsq plpsrlpkw segpcgspgs pgsniegfvc
 181 kfsfkgmcrp lalggpgqvt yttptqtss sleavpfasa anvacgegdk detqshyflc
 241 kekapdvfdw gssgplcvsp kygcfnfngg chqdcfeggd gsficgcrpg frliddlvtc

301 asrnpccssp crggatcvlg phgknytrc pqgyqidssq ldcvvdvdecq dspcaqecvn
 361 tpggfrcecw vgyepggpge gacqdvdeca lgrspcaqgc tntdgsfhcs ceegyvlage
 421 dgtqcqdvde cvpgggplcd slcftqgsf hcgcldpgwvl apngvsctmg pvsigppsgp
 481 pdeedkgeke gstvpraata sprtrgpgtp katpttsrps lssdapitsa plkmlapsgs
 5 541 pgvwrepsih hataasgpqe paggdssvat qnndgtgqk llifyilgtv vailllala
 601 lgllvykrk akreekkekk pqnaadsysw vperaesram enqysptpgt dc

SEQ ID NO: 5

10 O89103 Complement component C1q receptor precursor (Complement component
 1, q subcomponent, receptor 1) (C1qRp) (C1qR(p)) (C1q/MBL/SPA receptor) (CD93
 antigen) (Cell surface antigen AA4) (Lymphocyte antigen 68)
 gi|21541998|sp|O89103|CD93_MOUSE|21541998

15 source 1..644 /organism="Mus musculus" /db_xref="taxon:10090"
 gene 1..644 /gene="C1QR1" /note="CD93; C1QRP; LY68; AA4"
 Protein 1..644 /gene="C1QR1" /product="Complement component C1q receptor
 precursor"
 Region 1..22 /gene="C1QR1" /region_name="Signal" /note="POTENTIAL."
 Region 23..644 /gene="C1QR1" /region_name="Mature chain"
 20 /note="COMPLEMENT COMPONENT C1Q RECEPTOR."
 Region 23..572 /gene="C1QR1" /region_name="Domain" /note="EXTRACELLULAR
 (POTENTIAL)."
 Region 31..173 /gene="C1QR1" /region_name="Domain" /note="C-TYPE LECTIN."
 Site 102 /gene="C1QR1" /site_type="glycosylation" /note="N-LINKED (GLCNAC...)
 25 (POTENTIAL)."
 Region 257..298 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 1."
 Bond bond(261,272) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Bond bond(268,282) /gene="C1QR1" /bond_type="disulfide" /note="BY
 30 SIMILARITY."
 Bond bond(284,297) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Region 299..341 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 2."
 Bond bond(303,314) /gene="C1QR1" /bond_type="disulfide" /note="BY
 35 SIMILARITY."
 Bond bond(308,325) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Site 322 /gene="C1QR1" /site_type="glycosylation" /note="N-LINKED (GLCNAC...)
 (POTENTIAL)."
 40 Bond bond(327,340) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Region 342..381 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 3,
 CALCIUM-BINDING (POTENTIAL)."
 Bond bond(346,355) /gene="C1QR1" /bond_type="disulfide" /note="BY
 45 SIMILARITY."
 Bond bond(351,364) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Bond bond(366,380) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 50 Region 382..423 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 4,
 CALCIUM-BINDING (POTENTIAL)."
 Bond bond(386,397) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."

64

Bond bond(393,406) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Bond bond(408,422) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 5 Region 424..465 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 5,
 CALCIUM-BINDING (POTENTIAL)."
 Bond bond(428,440) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Bond bond(436,449) /gene="C1QR1" /bond_type="disulfide" /note="BY
 10 SIMILARITY."
 Bond bond(451,464) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Region 573..593 /gene="C1QR1" /region_name="Transmembrane region"
 /note="POTENTIAL."
 15 Region 594..644 /gene="C1QR1" /region_name="Domain" /note="CYTOPLASMIC
 (POTENTIAL)."
 ORIGIN 1 maistglfil lgllgqpwag. aaadsqavvc egtacytahw gklisaaeah rcnenggnla
 61 tvkseearh vqaltqllk tkapleakmg kfwiglqrek gnctyhdipm rgfswvggge
 121 dtaysnwyka sksscfkrk vsliildslt phpshlpkwh espcgtpeap gnsiegflck
 20 181 fnfkgmcrpl algpgprvty ttpfqtatss leavpfasva nvacgdeaks ethylcnek
 241 tpgifhwgss gplcvspkfg csfnnggcqq dcfeggdgsf rcgcrpgfrl lddlvtcasr
 301 npcssnpctg ggmchsvpls enyfcrcpsg yqldssqvhc vdidecqdsp caqdcvntlg
 361 sfhcecwwgy qpsgpkeeac edvdecaaan spcaqgcint dgsfycscke gyivsgedst
 421 qcedidecsd argnpcdslc fntdgsfrcg cppgwelapn gvfcsgrtvf selparppqk
 25 481 ednddrkest mpptempssp sgskdvsra qttglfvqsd iptasvplei eipsevsdvw
 541 felgtylptt sghskpthed svshsdtgd qnllfyilg tvvaisllv lalgiliyhk
 601 rrakkeike kkpqnaadsy swvperaesq apenqysptp gtgc

 SEQ ID NO: 6
 30 P09871 Complement C1s component precursor (C1 esterase)
 gi|115205|sp|P09871|C1S_HUMAN[115205]

 source 1..688 /organism="Homo sapiens" /db_xref="taxon:9606"
 gene 1..688 /gene="C1S"
 35 Protein 1..688 /gene="C1S" /product="Complement C1s component precursor"
 /EC_number="3.4.21.42"
 Region 1..15 /gene="C1S" /region_name="Signal"
 Region 16..437 /gene="C1S" /region_name="Mature chain" /note="COMPLEMENT
 C1S HEAVY CHAIN." Region 16..130 /gene="C1S" /region_name="Domain"
 40 /note="CUB 1."
 Bond bond(65,83) /gene="C1S" /bond_type="disulfide"
 Region 131..172 /gene="C1S" /region_name="Domain" /note="EGF-LIKE,
 CALCIUM-BINDING (POTENTIAL)."
 Bond bond(135,147) /gene="C1S" /bond_type="disulfide"
 45 Bond bond(143,156) /gene="C1S" /bond_type="disulfide"
 Site 149 /gene="C1S" /site_type="hydroxylation" /note="(PROBABLE)."
 Bond bond(158,171) /gene="C1S" /bond_type="disulfide"
 Site 174 /gene="C1S" /site_type="glycosylation" /note="N-LINKED (GLCNAC...)."
 Region 175..290 /gene="C1S" /region_name="Domain" /note="CUB 2."
 50 Bond bond(175,202) /gene="C1S" /bond_type="disulfide" Bond bond(234,251)
 /gene="C1S" /bond_type="disulfide" Region 293..355 /gene="C1S"
 /region_name="Domain" /note="SUSHI 1."
 Bond bond(294,341) /gene="C1S" /bond_type="disulfide"

Region 294 /gene="C1S" /region_name="Conflict" /note="C -> K (IN REF. 6)."
 Bond bond(321,354) /gene="C1S" /bond_type="disulfide"
 Region 358..422 /gene="C1S" /region_name="Domain" /note="SUSHI 2."
 Bond bond(359,403) /gene="C1S" /bond_type="disulfide"
 5 Bond bond(386,421) /gene="C1S" /bond_type="disulfide"
 Site 406 /gene="C1S" /site_type="glycosylation" /note="N-LINKED (GLCNAC...)."
 Bond bond(425,549) /gene="C1S" /bond_type="disulfide" /note="INTERCHAIN."
 Region 438..688 /gene="C1S" /region_name="Mature chain" /note="COMPLEMENT
 C1S LIGHT CHAIN." Region 438..688 /gene="C1S" /region_name="Domain"
 10 /note="SERINE PROTEASE."
 Site 475 /gene="C1S" /site_type="active" /note="CHARGE RELAY SYSTEM."
 Region 513 /gene="C1S" /region_name="Conflict" /note="G -> GG (IN REF. 5)."
 Site 529 /gene="C1S" /site_type="active" /note="CHARGE RELAY SYSTEM."
 Region 573 /gene="C1S" /region_name="Conflict" /note="T.-> A (IN REF. 7)."
 15 Bond bond(595,618) /gene="C1S" /bond_type="disulfide"
 Bond bond(628,659) /gene="C1S" /bond_type="disulfide"
 Site 632 /gene="C1S" /site_type="active" /note="CHARGE RELAY SYSTEM."
 Region 645..646 /gene="C1S" /region_name="Conflict" /note="TK -> GR (IN REF.
 7)."
 20 ORIGIN 1 mwcivilfll awvyaeptmy geilspnypq aypseveksw dievpegygi hlyfthldie
 61 lsencaydsv qiisgdteeg rlcqgrssnn phspiveefq vpynklqvif ksdfsneerf
 121 tgaayyvavt diiectdfvd vpcshfcnnf iggyfcscpp eyflhddmkn cgvnscgdvf
 181 taligeiasp nypkpypens rceyqirlek gqvvvflrr edfdveaads agnoldslvf
 241 vagdrqfgy cghgfppln ietksnaldi ifqtdltgqk kgwklyrhgd pmpcpkedtp
 25 301 nswwepakak yvfrdvvqit cldgfevveg rvgatsfyst cqsngkwsns klkcqpvdcg
 361 ipesiengkv edpesifgs vityceepy yymengggge yhcangswv nevlgpelpk
 421 cvpvcgvpre pfeekqriig gsdadiknfp wqvffdnppa ggaineiywv ltaahvvegn
 481 reptmyvgst svqtsrlaks kmllpehvf hpgwkllvpe egrtnfdndi alvrldpdk
 541 mgptvspicl pgtssdynlm dgdglisgw grtekdrav rikaarpva plrkckevkv
 30 601 ekptadaeay vftpnmicag gekgmdsckg dsggafavqd pndktkfyaa glvswgppcg
 661 tyglytrvkn yvdwimktmq enstpred

SEQ ID NO: 7

NP_036204 complement component 1, q subcomponent, receptor 1; complement
 35 component C1q receptor [Homo sapiens]
 gi|6912282|ref|NP_036204.1|[6912282]

source 1..652 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="20"
 /map="20p11.21"
 40 Protein 1..652 /product="complement component 1, q subcomponent, receptor 1"
 /note="complement component C1q receptor"
 Region 32..130 /region_name="smart00034, CLECT, C-type lectin (CTL) or
 carbohydrate-recognition domain (CRD); Many of these domains function as
 calcium-dependent carbohydrate binding modules"
 45 Region 47..128 /region_name="pfam00059, lectin_c, Lectin C-type domain. This
 family includes both long and short form C-type"
 Region 385..426 /region_name="smart00179, EGF_CA, Calcium-binding EGF-like
 domain"
 CDS 1..652 /gene="C1QR1" /coded_by="NM_012072.2:149..2107"
 50 /note="C1q/MBL/SPA receptor" /db_xref="LocusID:22918" /db_xref="MIM:120577"
 ORIGIN 1 matsmgllll lllltqpga gtgadteavv cvgtacytah sgklisaaeq nhcnqnggnl
 61 atvkskeaq hvqrvlaql rreaaltarm skfwiglqre kgkclpdlp lkgfswvggg
 121 edtpysnwhk elnsciskr cvslldlsq pllprlpkw segpcgspgs pgsniegfv

181 kfsfkgmcrp lalggpgqvt yttfqtss sleavpfasa anvacgegdk detqshyflc
 241 kekadvfdw gssgplcvsp kygcfnfngg chqdcfeggd gsflcgrpg frliddlvtc
 301 asmpcssp crgatcvg phgknytrc pqgyqldssq ldcvdvdecq dspcaqecvn
 361 tpggfrcecw vgyepggpge gacqdvdeca lgrspcagc tntdgsfhcs ceegyvlage
 421 dgtqcqdvde cvgpggplcd slcfnqtggsf hcgcldpgwvl apngvscimg pvsigppsgp
 481 pdeedkgeke gstvpraata sptrgpegtp katptsrps lssdapitsa plkmlapsgs
 541 sgvwrepsih hataasgpqe paggdssvat qnndgtgqk llifyilgtv vaillllala
 601 lgllyvrkrr akreekkekk pqnaadsysw vperaesram enqysptpgt dc

10 SEQ ID NO: 8
 NP_000233 soluble mannose-binding lectin precursor; mannose-binding lectin;
 mannose binding protein; Mannose-binding lectin 2, soluble (opsonic defect) [Homo
 sapiens]
 gi|4557739|ref|NP_000233.1|[4557739]

15 sig_peptide 1..20
 mat_peptide 21..248 /product="soluble mannose-binding lectin"
 variation 54 /allele="D" /allele="G" /db_xref="dbSNP:1800450"
 variation 57 /allele="E" /allele="G" /db_xref="dbSNP:1800451"
 20 Region 134..245 /region_name="smart00034, CLECT, C-type lectin (CTL) or
 carbohydrate-recognition domain (CRD); Many of these domains function as
 calcium-dependent carbohydrate binding modules"
 Region 144..246 /region_name="pfam00059, lectin_c, Lectin C-type domain. This
 family includes both long and short form C-type"
 25 CDS 1..248 /gene="MBL2" /coded_by="NM_000242.1:66..812"
 /db_xref="LocusID:4153" /db_xref="MIM:154545"
 ORIGIN 1 mslfplppll lsmvaasys etvtcedaqk tcpaviacss pgingfpgkd grdgtkgekg
 61 epqgqlrglq gppgklgppg nppsgspgp kgqkgdpgks pdgdsslaas erkalqtema
 121 rikkwitfsl gkqvgnkffl tngeimtfek vkalcvkfqa svatprnaae ngaiqnlike
 30 181 eafgitdek tegqfvdltg nrlytnwne gepnagsde dcvllknqg wndvpcastsh
 241 lavcefpj

SEQ ID NO: 9
 P11226 Mannose-binding protein C precursor (MBP-C) (MBP1) (Mannan-binding
 35 protein) (Mannose-binding lectin) gi|126676|sp|P11226|MABC_HUMAN[126676]

source 1..248 /organism="Homo sapiens" /db_xref="taxon:9606"
 gene 1..248 /gene="MBL2" /note="MBL"
 Protein 1..248 /gene="MBL2" /product="Mannose-binding protein C precursor"
 40 Region 1..20 /gene="MBL2" /region_name="Signal"
 Region 21..248 /gene="MBL2" /region_name="Mature chain" /note="MANNOSE-
 BINDING PROTEIN C." Region 21..41 /gene="MBL2" /region_name="Domain"
 /note="CYS-RICH."
 Region 24 /gene="MBL2" /region_name="Variant" /note="T -> A (IN CHINESE).
 45 /FTId=VAR_013294."
 Region 42..99 /gene="MBL2" /region_name="Domain" /note="COLLAGEN-LIKE."
 Site 47 /gene="MBL2" /site_type="hydroxylation"
 Region 52 /gene="MBL2" /region_name="Variant" /note="R -> C (IN 0.05% OF
 EUROPEAN AND AFRICAN POPULATIONS). /FTId=VAR_008543."
 50 Region 54 /gene="MBL2" /region_name="Variant" /note="G -> D (IN CAUCASIAN
 AND CHINESE POPULATIONS). /FTId=VAR_004182."
 Region 57 /gene="MBL2" /region_name="Variant" /note="G -> E (IN WEST
 AFRICAN POPULATION). /FTId=VAR_004183."

Site 73 /gene="MBL2" /site_type="hydroxylation"
 Site 79 /gene="MBL2" /site_type="hydroxylation"
 Site 82 /gene="MBL2" /site_type="hydroxylation"
 Site 88 /gene="MBL2" /site_type="hydroxylation"
 5 Region 109 /gene="MBL2" /region_name="Hydrogen bonded turn"
 Region 110..129 /gene="MBL2" /region_name="Helical region"
 Region 130 /gene="MBL2" /region_name="Hydrogen bonded turn"
 Region 132..134 /gene="MBL2" /region_name="Beta-strand region"
 Region 135..136 /gene="MBL2" /region_name="Hydrogen bonded turn"
 10 Region 137..147 /gene="MBL2" /region_name="Beta-strand region"
 Region 148..157 /gene="MBL2" /region_name="Helical region"
 Region 153..246 /gene="MBL2" /region_name="Domain" /note="C-TYPE LECTIN
 (SHORT FORM)."
 Bond bond(155,244) /gene="MBL2" /bond_type="disulfide"
 15 Region 158..159 /gene="MBL2" /region_name="Hydrogen bonded turn"
 Region 161..162 /gene="MBL2" /region_name="Beta-strand region"
 Region 168..177 /gene="MBL2" /region_name="Helical region"
 Region 182..187 /gene="MBL2" /region_name="Beta-strand region"
 Region 192..193 /gene="MBL2" /region_name="Hydrogen bonded turn"
 20 Region 196..197 /gene="MBL2" /region_name="Beta-strand region"
 Region 198..199 /gene="MBL2" /region_name="Hydrogen bonded turn"
 Region 202 /gene="MBL2" /region_name="Beta-strand region"
 Region 208 /gene="MBL2" /region_name="Beta-strand region"
 Region 210..211 /gene="MBL2" /region_name="Hydrogen bonded turn"
 25 Region 216..218 /gene="MBL2" /region_name="Helical region"
 Bond bond(222,236) /gene="MBL2" /bond_type="disulfide"
 Region 222..225 /gene="MBL2" /region_name="Beta-strand region"
 Region 227..228 /gene="MBL2" /region_name="Hydrogen bonded turn"
 Region 231..234 /gene="MBL2" /region_name="Beta-strand region"
 30 Region 236..237 /gene="MBL2" /region_name="Hydrogen bonded turn"
 Region 239..248 /gene="MBL2" /region_name="Beta-strand region"

ORIGIN 1 mslfplpII lsmvaasys etvtcedaqk tcpaviacss pgingfpgkd grdgtkgekq
 61 epqgqlrglq gppgklgppg nppgsgspgp kgqkgdpgks pdgdsslaas erkalqtema
 121 rikkwltfsl gkqvgnkffl tngeimtfek vkalcvkfqa svatprnaae ngaiqnlike
 181 eafllgitdek tegqfvdltg nrlytnwne gepnnagsde dcvlllknngq wndvpcastsh
 241 lavcefpI

SEQ ID NO: 10
 40 Q9ET61 Complement component C1q receptor precursor (Complement component
 1, q subcomponent, receptor 1) (C1qRp) (C1qR(p)) (C1q/MBL/SPA receptor) (CD93
 antigen) (Cell surface antigen AA4) gi|21541989|sp|Q9ET61|CD93_RAT[21541989]

source 1..643 /organism="Rattus norvegicus" /db_xref="taxon:10116"
 45 gene 1..643 /gene="C1QR1" /note="CD93; C1QRP2"
 Protein 1..643 /gene="C1QR1" /product="Complement component C1q receptor
 precursor"
 Region 1..23 /gene="C1QR1" /region_name="Signal" /note="POTENTIAL."
 Region 24..643 /gene="C1QR1" /region_name="Mature chain"
 50 /note="COMPLEMENT COMPONENT C1Q RECEPTOR."
 Region 24..571 /gene="C1QR1" /region_name="Domain" /note="EXTRACELLULAR
 (POTENTIAL)."
 Region 31..173 /gene="C1QR1" /region_name="Domain" /note="C-TYPE LECTIN."

Region 257..298 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 1."
 Bond bond(261,272) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 5 Bond bond(268,282) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Bond bond(284,297) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 10 Region 299..341 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 2."
 Bond bond(303,314) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Bond bond(308,325) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 15 Site 322 /gene="C1QR1" /site_type="glycosylation" /note="N-LINKED (GLCNAC...)
 (POTENTIAL)."
 Bond bond(327,340) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Region 342..381 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 3,
 CALCIUM-BINDING (POTENTIAL)."
 20 Bond bond(346,355) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Bond bond(351,364) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Bond bond(366,380) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 25 Region 382..423 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 4,
 CALCIUM-BINDING (POTENTIAL)."
 Bond bond(386,397) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 30 Bond bond(393,406) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Bond bond(408,422) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Region 417 /gene="C1QR1" /region_name="Conflict" /note="E -> K (IN REF. 2)."
 Region 424..462 /gene="C1QR1" /region_name="Domain" /note="EGF-LIKE 5,
 35 CALCIUM-BINDING (POTENTIAL)."
 Bond bond(428,437) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Bond bond(433,446) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 40 Bond bond(448,461) /gene="C1QR1" /bond_type="disulfide" /note="BY
 SIMILARITY."
 Site 498 /gene="C1QR1" /site_type="glycosylation" /note="N-LINKED (GLCNAC...)
 (POTENTIAL)."
 45 Region 572..592 /gene="C1QR1" /region_name="Transmembrane region"
 /note="POTENTIAL."
 Region 593..643 /gene="C1QR1" /region_name="Domain" /note="CYTOPLASMIC
 (POTENTIAL)."
 50 ORIGIN 1 mvtstgllll lglgqlwag aaadseavvc egtacytahw.gklsaaeahq rcnenggnla
 61 .tvkseearh vqealaqlk tkapsetkig kfwiglqrek gkctyhdlpm kgfswvggge
 121 dttysnwyka skssciskrc vsliildslk phpshlpkwh espcgtpdap gnsiegflick
 181 fnfkgmcspl algpggqlty ttpfqattss lkavpfasva nvvcgdeaes ktnyylcket
 241 tagvfhwgss gplcvspkfg csfnnggcq qdcfeggdgsf rcgcrpgfrl lddlvtcasr

301 npcssnptcg ggmchsvpls enytchcprg yqldssqvhc vdidecedsp cdqecintpg
 361 gfhcecwvgy qssgskeeac edvdectaaay spcaqgctnt dgsfycscke gyimsgedst
 421 qcedideclg npcdtlcint dgsfrcgcpa gfelapngvs ctrgsmfsej parppqkedk
 481 gdgkestvpl tempgslns kdvsnraqtt dlsiqsdsst asvpleievs seasdvwldl
 5 541 gtylpttsg hsqphedsvp ahsdsdtdgq kllfyilgt vvaisllal algliylkr
 601 kakkeeikek kaqnaadsys wiperaesra penqysptpg tdc

SEQ ID NO: 11

NP_006601 mannan-binding lectin serine protease 2, isoform 1 precursor; MBL-associated plasma protein of 19 kD; small MBL-associated protein [Homo sapiens] gi|21264363|ref|NP_006601.2||21264363]
 sig_peptide 1..15
 mat_peptide 16..444 /product="mannan-binding lectin serine protease 2, isoform 1, chain A"
 15 Region 28..136 /region_name="Domain first found in C1r, C1s, uEGF, and bone morphogenetic protein." /note="CUB" /db_xref="CDD:smart00042"
 Region 28..134 /region_name="CUB domain" /note="CUB"
 /db_xref="CDD:pfam00431"
 Region 138..180 /region_name="Calcium-binding EGF-like domain"
 20 /note="EGF_CA" /db_xref="CDD:smart00179"
 variation 155 /allele="H" /allele="R" /db_xref="dbSNP:2273343"
 Region 184..295 /region_name="Domain first found in C1r, C1s, uEGF, and bone morphogenetic protein." /note="CUB" /db_xref="CDD:smart00042"
 Region 184..293 /region_name="CUB domain" /note="CUB"
 25 /db_xref="CDD:pfam00431"
 Region 300..361 /region_name="Domain abundant in complement control proteins" /note="CCP" /db_xref="CDD:smart00032"
 Region 300..361 /region_name="Sushi domain (SCR repeat)" /note="sushi"
 /db_xref="CDD:pfam00084"
 30 Region 366..430 /region_name="Domain abundant in complement control proteins" /note="CCP" /db_xref="CDD:smart00032"
 Region 366..430 /region_name="Sushi domain (SCR repeat)" /note="sushi"
 /db_xref="CDD:pfam00084"
 variation 377 /allele="A" /allele="V" /db_xref="dbSNP:2273346"
 35 Region 444..679 /region_name="Trypsin-like serine protease" /note="Tryp_SPc" /db_xref="CDD:smart00020" mat_peptide 445..686 /product="mannan-binding lectin serine protease 2, isoform 1, chain B"
 Region 445..679 /region_name="Trypsin" /note="trypsin" /db_xref="CDD:pfam00089"
 40 CDS 1..686 /gene="MASP2" /coded_by="NM_006610.2:22..2082" /db_xref="LocusID:10747" /db_xref="MIM:605102"

ORIGIN 1 mrltlgl cgsvatplgp kwpepvfgrl aspgfpgeya ndqerrwtl appgyrlrl
 61 fthfdlelsh lceydfvkl sgakvlatlc gqestdtera pgkdtfyslg sslditfrsd
 45 121 ysnekpftgf eafyaaedid ecqvapegap tcdhhchnhl ggfyccrag yvlhmkrct
 181 salcsgqvft qrsgelsspe yprpypklss ctysisleeg fsvldfves fdvethpetl
 241 cpydflikiqt dreehgpfcg ktlphriek sntvtitfv desgdhtgwk ihtystaqpc
 301 pypmappngh vspvqakyil kdsfsifcet gyellqghlp lksftavcck dgswdrpmpa
 361 csivdcgppd dlpsgrveyi tpggvtyka viqysceetf ytmkvndgky vceadgfwts
 50 421 skgekslpvc epvcglart tggriggqk akpgdfpwqv lilggtaag allydnwvlt
 481 aahavyeqkh dasaldirng tikrlsphyt qawseavfih egythdagfd ndialiklhn
 541 kvvinsnitp iclprkeaes fmrtddigta sgwgltrgf larnlmyvdi pivdhqkcta
 601 ayakppyprg svtanmlcag lesggkdscr gdsggalvfl dseterwfv givswgsmnc

661 geagqygvyt kvinyipwie niisdf

SEQ ID NO: 12

NP_631947 mannan-binding lectin serine protease 2, isoform 2 precursor; MBL-associated plasma protein of 19 kD; small MBL-associated protein [Homo sapiens] gi|21264361|ref|NP_631947.1|[21264361]
 sig_peptide 1..15
 mat_peptide 16..185 /product="mannan-binding lectin serine protease 2, isoform 2"
 Region 28..136 /region_name="Domain first found in C1r, C1s, uEGF, and bone morphogenetic protein." /note="CUB" /db_xref="CDD:smart00042"
 Region 28..134 /region_name="CUB domain" /note="CUB" /db_xref="CDD:pfam00431"
 Region 138..180 /region_name="Calcium-binding EGF-like domain" /note="EGF_CA" /db_xref="CDD:smart00179"
 variation 155 /allele="H" /allele="R" /db_xref="dbSNP:2273343"
 CDS 1..185 /gene="MASP2" /coded_by="NM_139208.1:22..579" /db_xref="LocusID:10747" /db_xref="MIM:605102"
 ORIGIN 1 mrltlgl cgsvatplgp kwpepvfgrl aspgfpgeya ndqerrwtlt appgyrlrl
 61 fthfdlelsh lceydfvklis sgakvlatlc gqestdtera pgkdtfyslg sslditfrsd
 121 ysnekpftgf eafyaaedid ecqvapgeap tcdhhchnhl ggfycscrag yvlhrnkrtc
 181 seqsl

SEQ ID NO: 13

NP_624302 mannan-binding lectin serine protease 1, isoform 2, precursor; protease, serine, 5 (mannose-binding protein-associated); mannan-binding lectin serine protease-1; Ra-reactive factor serine protease p100 [Homo sapiens] gi|21264359|ref|NP_624302.1|[21264359]
 sig_peptide 1..19
 mat_peptide 20..445 /product="mannan-binding lectin serine protease 1, isoform 2, chain A"
 variation 21 /allele="I" /allele="T" /db_xref="dbSNP:1062049"
 Region 23..138 /region_name="Domain first found in C1r, C1s, uEGF, and bone morphogenetic protein." /note="CUB" /db_xref="CDD:smart00042"
 Region 23..135 /region_name="CUB domain" /note="CUB" /db_xref="CDD:pfam00431"
 Region 139..181 /region_name="Calcium-binding EGF-like domain" /note="EGF_CA" /db_xref="CDD:smart00179"
 Region 185..294 /region_name="CUB domain" /note="CUB" /db_xref="CDD:pfam00431"
 Region 190..296 /region_name="Domain first found in C1r, C1s, uEGF, and bone morphogenetic protein." /note="CUB" /db_xref="CDD:smart00042"
 variation 235 /allele="Q" /allele="E" /db_xref="dbSNP:3203210"
 variation 258 /allele="P" /allele="A" /db_xref="dbSNP:866085"
 Region 301..362 /region_name="Domain abundant in complement control proteins" /note="CCP" /db_xref="CDD:smart00032"
 Region 301..362 /region_name="Sushi domain (SCR repeat)" /note="sushi" /db_xref="CDD:pfam00084"
 Region 367..432 /region_name="Domain abundant in complement control proteins" /note="CCP" /db_xref="CDD:smart00032"
 Region 367..432 /region_name="Sushi domain (SCR repeat)" /note="sushi" /db_xref="CDD:pfam00084" mat_peptide 446..728 /product="mannan-binding lectin serine protease 1, isoform 2, chain B"

- Region 449..711 /region_name="Trypsin-like serine protease" /note="Tryp_SPC"
 /db_xref="CDD:smart00020" Region 450..711 /region_name="Trypsin"
 /note="trypsin" /db_xref="CDD:pfam00089"
 variation 616 /allele="A" /allele="V" /db_xref="dbSNP:2461280"
- 5 Region 661..703 /region_name="Immunoglobulin A1 protease" /note="IGA1"
 /db_xref="CDD:pfam02395"
 CDS 1..728 /gene="MASP1" /coded_by="NM_139125.1:51..2237"
 /db_xref="LocusID:5648" /db_xref="MIM:600521"
- 10 ORIGIN 1 mrwillyal cfslskasah tvelnnmfgq iqspgypdsy psdsevtwni tvpdgfrik
 61 yfmhfnless ylceydykv etedqvlaf cgrettdteq tpgqevlsp gsfnstfrs
 121 dfsneerftg fdahymavdv deckeredee lscdhychny iggyycscr gylhtdnrt
 181 crvecsdnlf tqrtgvitp dfpnypkss eelytielee gfmvnlqfed ifdiedhpev
 241 pcpydyikik vgpklvgpfc gekapepist qshsvliifh sdnsgenrgw rlsyraagne
 15 301 cpelqppvhg kiepsqakyf fkdqvlvscd tgykvikdnv emdtfqiecl kdgtswnkip
 361 tckivdcrap gelehglitf strnlttyk seikyscqp yykmlhntg iytsaagvw
 421 mnkvlgrslp tclpecgqps rslpslvkri iggrnaepgl fpwqalivve dtsrvpndkw
 481 fgsgallsas wлтаahvlr sqrrdtvip vskehvtvyl glhdvrdksg avnssaarvv
 541 lhpdfniqny nhdialvqlq epvplgphvm pvcplrepe gpaphmlglv agwgisnprv
 20 601 tvdeiissgt rtlsdvlqv klpvvpahac ktsyesrsgn ysvenmfca gyyeggkdtc
 661 lgdsggafvi fddlsqrwv qglvswggpe ecgskqvvgv ytkvsnyvdw vweqmglpqs
 721 vvepqver
- SEQ ID NO: 14
- 25 NP_001870 mannan-binding lectin serine protease 1, isoform 1, precursor;
 protease, serine, 5 (mannose-binding protein-associated); manan-binding lectin
 serine protease-1; Ra-reactive factor serine protease p100 [Homo sapiens]
 gi|21264357|ref|NP_001870.3||21264357]
- 30 sig_peptide 1..19
 mat_peptide 20..448 /product="mannan-binding lectin serine protease 1, isoform 1,
 chain A"
 variation 21 /allele="I" /allele="T" /db_xref="dbSNP:1062049"
- 35 Region 23..138 /region_name="Domain first found in C1r, C1s, uEGF, and bone
 morphogenetic protein." /note="CUB" /db_xref="CDD:smart00042"
 Region 23..135 /region_name="CUB domain" /note="CUB"
 /db_xref="CDD:pfam00431"
 Region 139..181 /region_name="Calcium-binding EGF-like domain"
 /note="EGF_CA" /db_xref="CDD:smart00179"
- 40 Region 185..294 /region_name="CUB domain" /note="CUB"
 /db_xref="CDD:pfam00431"
 Region 190..296 /region_name="Domain first found in C1r, C1s, uEGF, and bone
 morphogenetic protein." /note="CUB" /db_xref="CDD:smart00042"
 variation 235 /allele="Q" /allele="E" /db_xref="dbSNP:3203210"
- 45 variation 258 /allele="P" /allele="A" /db_xref="dbSNP:866085"
 Region 301..362 /region_name="Domain abundant in complement control proteins"
 /note="CCP" /db_xref="CDD:smart00032"
 Region 301..362 /region_name="Sushi domain (SCR repeat)" /note="sushi"
 /db_xref="CDD:pfam00084"
- 50 Region 367..432 /region_name="Domain abundant in complement control proteins"
 /note="CCP" /db_xref="CDD:smart00032"
 Region 367..432 /region_name="Sushi domain (SCR repeat)" /note="sushi"
 /db_xref="CDD:pfam00084"

72

Region 448..691 /region_name="Trypsin-like serine protease" /note="Tryp_SPC"
/db_xref="CDD:smart00020" mat_peptide 449..699 /product="mannan-binding lectin
serine protease 1, isoform 1, chain B"

Region 449..691 /region_name="Trypsin" /note="trypsin"

/db_xref="CDD:pfam00089"

Region 644..675 /region_name="Immunoglobulin A1 protease" /note="IGA1"

/db_xref="CDD:pfam02395"

CDS 1..699 /gene="MASP1" /coded_by="NM_001879.3:51..2150"

/db_xref="LocusID:5648" /db_xref="MIM:600521"

ORIGIN 1 mrwillyyal cfslskasah tvelnnmfgq iqspgypdsy psdsevtwni tvpdgfrkl
61 yfmhfnless ylceydykv etedqvlaf cgrettdteq tpgqevvlsps gsfmsitfrs
121 dfsneerftg fdahymavdv deckeredee lscdhychny iggyycscr fgyilhtdnrt
181 crvecsdlf tqrtrgvtsp dfpnypkss eclytielee gfmvnlqfed ifdiedhpev
241 pcpydykik vgpklvgpfc gekapepist qshsvlilfh sdnsgenrgw rlsyraagne
301 cpelqppvhg kiepsqakyf fkdqvlvsd tgykvlkdnv emdtfqiecl kdgtswnkip
361 tckivdcrap gelehglitf strnnlttyk seikyscqp yykmlnnntg iytcsaaggw
421 mnkvlgslp tclpvcglpk fsrklmarif ngrpaqkgtt pwiamlshln gqpfcggsll
481 gsswivtaah clhqsldped ptlrdsdls psdfkiilgk hwrlrsdne qhlgvkhltl
541 hpqydpntfe ndvalvelle spvlnafvmp iclpegpqqe gamvivsgwg kqflqrpet
601 lmeieipivd hstcqkayap lkkkvtrdmi cagekeggkd acagdsaggpm vtlrnergqw
661 ylygtvswgd dcgkkdrygv ysiihnhkdw iqrvtgvn

SEQ ID NO: 15

XP_122683 similar to mannose binding lectin, liver (A) [Mus musculus]

gi|20872845|ref|XP_122683.1|[20872845]

source 1..239 /organism="Mus musculus" /strain="C57BL/6J"

/db_xref="taxon:10090" /chromosome="14"

Protein 1..239 /product="similar to mannose binding lectin, liver (A)"

Region 126..236 /region_name="C-type lectin (CTL) or carbohydrate-recognition
domain (CRD)" /note="CLECT" /db_xref="CDD:smart00034"

Region 135..237 /region_name="Lectin C-type domain" /note="lectin_c"
/db_xref="CDD:pfam00059"

CDS 1..239 /gene="Mbl1" /coded_by="XM_122683.1:10..729"

/db_xref="LocusID:17194" /db_xref="MGD:96923"

ORIGIN 1 mlllpplpvl lcvsvsyssg sqtcedtlkt csviacgrdg rdgpkgekge pgqglrglqg
61 ppgklgppgs vsgpgspgpk gqkgdhgdnr xxxxxxxxxx xxxxxxxxxx xxxxxxxhafs
121 mgkksqkklf vtnhekmpfs kvkslctelq gtvaipnae enkaiqevat giaflgitde
181 ategqfmyvt ggrltsynwk kdepnnhsg edcvildng lwndiscqas fkavcefa

SEQ ID NO:16

AAM21196 C-type mannose-binding lectin [Oncorhynchus mykiss]

gi|20385163|gb|AAM21196.1|AF363271_1[20385163]

source 1..185 /organism="Oncorhynchus mykiss" /db_xref="taxon:8022"

Protein 1..185 /product="C-type mannose-binding lectin"

CDS 1..185 /gene="MBL" /coded_by="AF363271.1:25..582"

ORIGIN 1 meklaiilll sasiaigdan ltqllglepl lktkveqtp eaqveavqeg ikegscpsdw

61 ytygshckf vsiqqsfdvs eqnclalgg lasvhsley qfmqaltkda nghlhwlg

121 gfdaikegtw mwsdgsrfdy tnwdtdepnn agegedclhm naasaklwd vpcewkfasl

181 csrrm

SEQ ID NO: 17

AAD45377 mannose-binding lectin [Sus scrofa]
gi|5566370|gb|AAD45377.1|AF164576_1|5566370]

5 source 1..240 /organism="Sus scrofa" /db_xref="taxon:9823" /tissue_type="liver"
Protein 1..240 /product="mannose-binding lectin"
CDS 1..240 /coded_by="AF164576.1:1..723"

10 ORIGIN 1 msifpslhl llivmtasht etencediqn tclviscdsp ginglpgkdg ldgakekge
61 pgqglglqg lpgmvgpqs pgipglpglk gqkgdsgidp gnslnlrse ldnkkwlif
121 aqgkqvqkkl yltngkkmsf ngvkalcaqf qasvatptns renqaiqela gteafgitd
181 eyteggqfvdI tgrvryqnw ndgepnads aehcveilkd gkwndifcss qlsavcefpA

SEQ ID NO: 18

15 NP_034905 mannose binding lectin, liver (A) [Mus musculus]
gi|6754654|ref|NP_034905.1|[6754654]

source 1..239 /organism="Mus musculus" /db_xref="taxon:10090"
/chromosome="14" /map="14 15.0 cM"
20 Protein 1..239 /product="mannose binding lectin, liver (A)"
misc_feature 19..239 /partial /note="mature protein based on homology to rat MPB-A"
Region 126..236 /region_name="C-type lectin (CTL) or carbohydrate-recognition domain (CRD)" /note="CLECT" /db_xref="CDD:smart00034"
25 Region 135..237 /region_name="Lectin C-type domain" /note="lectin_c"
/db_xref="CDD:pfam00059"
CDS 1..239 /gene="Mbl1" /coded_by="NM_010775.1:121..840"
/db_xref="LocusID:17194" /db_xref="MGD:96923"
ORIGIN 1 mlllpplpvl lcvsvsssg sqtcedtikt csiacgrdg rdgpkgekge pgqglrglqg
30 61 ppgklgppgs vgspspgpk gqkgdhgdnr aieeklanme aeirilkskl qltnklhafs
121 mgkksqgklf vtnhekmpfs kvksictelq gtvaipnae enkaiqevaf giaflgitde
181 ateqqfmyvt ggrltsnwk kdepnnhsg edcvilndng lwndiscqas fkavcefpA

SEQ ID NO: 19

35 NP_034906 mannose binding lectin, serum (C) [Mus musculus]
gi|6754656|ref|NP_034906.1|[6754656]

source 1..244 /organism="Mus musculus" /strain="BALB/c"
/sub_species="domesticus" /db_xref="taxon:10090" /chromosome="19" /map="19
40 25.0 cM" /clone="a10" /tissue_type="liver" /clone_lib="lambda gt10"
Protein 1..244 /product="mannose binding lectin, serum (C)"
sig_peptide 1..18
Region 120..241 /region_name="C-type lectin (CTL) or carbohydrate-recognition domain (CRD)" /note="CLECT" /db_xref="CDD:smart00034"
45 Region 140..242 /region_name="Lectin C-type domain" /note="lectin_c"
/db_xref="CDD:pfam00059"
CDS 1..244 /gene="Mbl2" /coded_by="NM_010776.1:177..911"
/note="polysaccharide-binding component of RaRF; sequence similarity to mannose-binding proteins" /db_xref="LocusID:17195" /db_xref="MGD:96924"
50 ORIGIN 1 msiftsflil cvvtvyaet ltegvqnsdp vvtcsspgln gfpkgdgrdg akgekgepgg
61 glrglqpppg kvgtptgpgn pglkgavgpk gdrgrdraefd tseidseiaa lrselraln
121 wvflsisekv gkkyfvssvk kmsldrvkal csefqgsvat prnaeensa qkvakdiayl

181 gitdvrvs fedlignrvr ytnwndgepn ntgdgedcvv ilgngkwndv pcsdsflaic
241 efsd

SEQ ID NO: 20

5 AAL14428 dendritic cell-specific ICAM-3 grabbing nonintegrin [Macaca nemestrina]
gi|16118455|gb|AAL14428.1|AF343727_1[16118455]

source 1..381 /organism="Macaca nemestrina" /db_xref="taxon:9545"
/cell_type="peripheral blood-derived dendritic cells"

10 Protein 1..381 /product="dendritic cell-specific ICAM-3 grabbing nonintegrin"
/name="membrane-associated mannose binding lectin"

CDS 1..381 /coded_by="AF343727.1:1..1146" /note="DC-SIGN"

ORIGIN 1 msdskepqli qldleeeql ggvgfrqtrg ykslagclgh gplvlqlsf tllagllvqv

15 61 skvpsslsqg qskqdaiyqn ltqlkvavse lsekskqei yqeltrkaa vgelpekskq
121 qeiyeeltri raavgelpek sklqeiyeql trkaavgel pekskqeiye qelsrlkaav
181 gdlpekskqq eiyqklitqlk aavdglpdrs kqeiyeqli qlkaaverlc hpcpwewtff
241 qgncyfmsns qmwhdsita cqvgaqlvv ksaaeqnfl qlqssrsnrf twmglsdlnh
301 egtwqwwdgs plpsfkqyw nkgepnvge edcaefsgng wdddkcnlak

20 fwickksaas

361 csgdeerlls paptpnppp a

SEQ ID NO: 21

25 AAF63470 mannose binding-like lectin precursor [Carassius auratus]
gi|7542474|gb|AAF63470.1|AF227739_1[7542474]

source 1..246 /organism="Carassius auratus" /db_xref="taxon:7957"
/tissue_type="liver"

Protein <1..246 /product="mannose binding-like lectin precursor" /name="collectin"
sig_peptide <1..13

30 Region 14..25 /region_name="N-terminal segment"

Region 26..93 /region_name="collagen-like structure"

Region 60..63 /region_name="break in collagen structure" Region 94..124
/region_name="neck region"

Region 125..246 /region_name="carbohydrate recognition domain" /note="CRD"

35 CDS 1..246 /gene="MBL" /coded_by="AF227739.1:<1..742" /note="collectin with
structural homology to mannose-binding lectin but with a predicted carbohydrate
specificity for galactose"

ORIGIN 1 llllqfalql ldgaepqnl ncpayggvpgt pghnglpgrd grdgkdgaig pkgekgesgv

40 61 svqgppgkag ppgtagekge rgpsgpqgsp gsesvlesk seiqqlkaki atfekvssvc
121 hfrkvqgkyy itdgvgnfd qglksmefg gtmvsprtsa enqalilvv ssglgskkpy
181 igvtdrkteg qfvdteqql tftnwpgqp ddykgldcg viedtglwdd ggcdgdirpm
241 ceidik

45 SEQ ID NO: 22

AAF63469 mannose binding-like lectin precursor [Danio rerio]
gi|7542472|gb|AAF63469.1|AF227738_1[7542472]

sig_peptide 1..23

mat_peptide 24..251 /product="mannose binding-like lectin"

50 Region 24..36 /region_name="N-terminal segment"

Region 37..101 /region_name="collagen-like structure"

Region 71..74 /region_name="break in collagen structure"

Region 102..132 /region_name="neck region"

Region 133..251 /region_name="carbohydrate recognition domain" /note="CRD"
 CDS 1..251 /gene="mbi" /coded_by="AF227738.1:68..823" /note="collectin with
 structural homology to mannose-binding lectin but with a predicted carbohydrate
 specificity for galactose"

5 ORIGIN 1 mallklflga lllqlvlql magaadpqsI ncpayagvpg tpghnglpgr dgrvgrdgan
 61 gpkgekgepg vnvqgppgka gppgpagakg ergpsglpgq dcmsdslkse lqklsdial
 121 iekvvnfktf kkvgqkyvt ddveetfdkg mqycssngga lvprtlelen allkvfvssa
 181 fkrfirtid rekegefvdI drkklftnw gpnqpdnykg aqdcgaiads glwddvscds
 241 lypiiceiei k

10

SEQ ID NO: 23

AAF63468 mannose binding-like lectin precursor [Cyprinus carpio]
 gi|7542470|gb|AAF63468.1|AF227737_1[7542470]

15

sig_peptide 1..23

mat_peptide 24..256 /product="mannose binding-like lectin"

Region 24..35 /region_name="N-terminal segment"

Region 36..103 /region_name="collagen-like structure"

Region 70..73 /region_name="break in collagen structure"

20

Region 104..134 /region_name="neck region"

Region 135..256 /region_name="carbohydrate recognition domain" /note="CRD"
 CDS 1..256 /gene="MBL" /coded_by="AF227737.1:67..837" /note="collectin with
 structural homology to mannose-binding lectin but with a predicted carbohydrate
 specificity for galactose"

25

ORIGIN 1 malfkflgt lllqfalql ldgaepqnln cpayggvpgt pghnglpgrd grdgkdgaig
 61 pkgekgesgv svqgppgkag ppgpagekge rgptgsqgsp gsesvleslk seiqqkaki
 121 atfekvasvg hfrqvgqkyi itdgvggtfd qglkfckdfg gtmvfprtsa enqallklvv
 181 ssglsskppy igvtdreteg rfvntegkql tftnwpggqp ddykgldcgc viedsglwd
 241 gscgdirpim ceidnk

30

SEQ ID NO: 24

AAF21018 mannose-binding lectin 2 [Sus scrofa]

gi|6644342|gb|AAF21018.1|AF208528_1[6644342]

35

source 1..31 /organism="Sus scrofa" /db_xref="taxon:9823" /chromosome="14"
 /map="between S0007 and Sw210" Protein <1..>31 /product="mannose-binding
 lectin 2" /name="MBL2"

CDS 1..31 /gene="MBL2"

/coded_by="join(AF208528.1:<1..25,AF208528.1:703..>771)"

40

ORIGIN 1 tkgekgepgp gfrgsqgppg kmgppgnige t

SEQ ID NO: 25

AAK30298 mannose-binding lectin precursor protein [Gallus gallus]

gi|13561409|gb|AAK30298.1|[13561409]

45

sig_peptide 1..21

mat_peptide 22..254 /product="mannose-binding lectin protein"

Region 22..46 /region_name="N-terminal segment"

Region 47..102 /region_name="collagen-like"

50

Region 66 /region_name="break in collagen-like structure"

Region 103..139 /region_name="neck region"

Region 140..254 /region_name="carbohydrate recognition domain; CRD"

CDS 1..254 /coded_by="AF231714.1:242..1006"

ORIGIN 1 mtlqpfsal llclslmmt sltttdkpee kmyscpilqc sapavnglpg rdgrdgpkge
 61 kgdpgeglrg lqglpgkagp qglkgevpgq gekgqkgerg ivvtdldhrq itdleakirv
 121 leddlsrykk alsikdvvnv gkkmfvstgk kynfekgksl cakagsvlas prneaental
 181 kdliidpssqa yigisdaqte grfmylsggp ltysnwkpgge pnnhknedca viedsgkwnd
 241 ldcnsnifi icel

SEQ ID NO: 26

LNMSMC mannose-binding lectin C precursor – mouse
 gi|7428747|pir||LNMSMC[7428747]

FEATURES Location/Qualifiers source 1..244 /organism="Mus musculus"
 /db_xref="taxon:10090"
 Protein 1..244 /product="mannose-binding lectin C precursor" /note="Ra-reactive
 factor P28a"
 Region 1..18 /region_name="domain" /note="signal sequence"
 Region 19..244 /region_name="product" /note="mannose-binding lectin C"
 Bond bond(29) /bond_type="disulfide" /note="interchain"
 Bond bond(34) /bond_type="disulfide" /note="interchain"
 Region 38..94 /region_name="region" /note="collagen-like"
 Site 69 /site_type="modified" /note="4-hydroxyproline (Pro)"
 Region 124..240 /region_name="domain" /note="C-type lectin homology #label
 LCH"

ORIGIN 1 msiftsfill cvvtvyaet ltegvqnsdp vvtcsspgln gfpkgdgrdg akgekgepgg
 61 glrglqgppg kvgptgppgn pglkgavgpk gdrgrdaefd tseidseiaa lrselralrn
 121 wvlfslsekv gkkyfvssvk kmsldrvkal csefggsvat pmaeensa qkvakdiayl
 181 gitdvrvegs fedltgnrvr ytnwndgepn ntgdgedcvv ilgngkwndv pcsdsflaic
 241 efsd

SEQ ID NO: 27

LNMSMA mannose-binding lectin A precursor – mouse
 gi|625320|pir||LNMSMA[625320]

FEATURES Location/Qualifiers source 1..239 /organism="Mus musculus"
 /db_xref="taxon:10090"
 Protein 1..239 /product="mannose-binding lectin A precursor" /note="Ra-reactive
 factor P28b; serum mannan-binding protein"
 Region 1..17 /region_name="domain" /note="signal sequence"
 Region 18..238 /region_name="product" /note="mannose-binding lectin A"
 Region 36..88 /region_name="region" /note="collagen-like"
 Region 119..235 /region_name="domain" /note="C-type lectin homology #label
 LCH"

ORIGIN 1 mlllpilpvl lcvsvsssg sqtcedtlkt csviacgrdg rdgpkgekge pgqglrglqg
 61 ppgklgppgs vsgpgspgpk gqkgdhgdnr aieeklanme aeirilkskl qltnklhafs
 121 mgkksqgklf vtnhekmfks kvkslctelq gtvaipmae enkaieqvai giagfigtde
 181 ategqfmyvt ggrltysnwk kdepnnhgsg edcvildng lwndiscqas fkavcefp

SEQ ID NO: 28

LNRTMA mannose-binding lectin A precursor – rat gi|71975|pir||LNRTMA[71975]

FEATURES Location/Qualifiers source 1..238 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"

Protein 1..238 /product="mannose-binding lectin A precursor" /note="serum
mannan-binding protein"
Region 1..17 /region_name="domain" /note="signal sequence"
Region 18..238 /region_name="product" /note="mannose-binding lectin A"
5 Region 36..88 /region_name="region" /note="collagen-like"
Site 61 /site_type="modified" /note="4-hydroxyproline (Pro)"
Site 67 /site_type="modified" /note="4-hydroxyproline (Pro)"
Site 73 /site_type="modified" /note="4-hydroxyproline (Pro)"
Site 79 /site_type="modified" /note="lysine derivative (Lys) (probably 5-
10 hydroxylysine)"
Site 82 /site_type="modified" /note="lysine derivative (Lys) (probably 5-
hydroxylysine)"
Region 85..87 /region_name="region" /note="cell attachment (R-G-D) motif"
Region 118..234 /region_name="domain" /note="C-type lectin homology #label
15 LCH"

ORIGIN 1 mllpllvll cvsvsssgs qtceetlktc sviacgrdgr dgpkgekgep gqglrglqgp
61 pgklgppgsv gapgsqgpkq kqgdrgdsra ievklanmea eintlkskle ltnklhafsm
121 gkksqgkffv tnhermpfsk vkalcseirg tvaipmaee nkaiqevakt saflgitdev
20 181 tegqfmyvtg grltylnwkk depndhgsge dcvtivdngl wndiscqash tavcefp

SEQ ID NO: 29
LNRTMC mannanose-binding lectin C precursor – rat gi|71974|pir||LNRTMC[71974]
FEATURES Location/Qualifiers source 1..244 /organism="Rattus norvegicus"
25 /db_xref="taxon:10116"
Protein 1..244 /product="mannose-binding lectin C precursor"
Region 1..18 /region_name="domain" /note="signal sequence"
Region 19..244 /region_name="product" /note="mannose-binding lectin C"
Bond bond(29) /bond_type="disulfide" /note="interchain"
30 Bond bond(34) /bond_type="disulfide" /note="interchain"
Region 38..94 /region_name="region" /note="collagen-like"
Site 69 /site_type="modified" /note="4-hydroxyproline (Pro)"
Region 124..240 /region_name="domain" /note="C-type lectin homology #label
LCH"
35 ORIGIN 1 mslftsfill cvltavyaet ltegaqsscp viacsspgin gfpkgdghdg akgekgepgq
61 glrglqgppg kvgapppgn pgskgatgpk gdrgevefd ttnidleiaa lrselramrk
121 wvllmseniv gkkyfmssvr mmlnrakal cselqgtvat prnaeenrai qnvakdvafll
181 gitdqrtenv fedltgnrvr ytnwnegepn nvsgsencvv lltngkwndv pcsdsflvvc
241 efsd
40

SEQ ID NO: 30
LNHUMC mannanose-binding lectin precursor – human gi|71973|pir||LNHUMC[71973]
FEATURES Location/Qualifiers source 1..248 /organism="Homo sapiens"
/db_xref="taxon:9606"
45 Protein 1..248 /product="mannose-binding lectin precursor" /note="mannan-binding
protein"
Region 1..20 /region_name="domain" /note="signal sequence"
Region 21..248 /region_name="product" /note="mannose-binding lectin"
Region 42..99 /region_name="region" /note="collagen-like"
50 Site 47 /site_type="modified" /note="4-hydroxyproline (Pro) (partial)"
Site 73 /site_type="modified" /note="4-hydroxyproline (Pro) (partial)"
Site 79 /site_type="modified" /note="4-hydroxyproline (Pro) (partial)"
Site 82 /site_type="modified" /note="4-hydroxyproline (Pro) (partial)"

Site 88 /site_type="modified" /note="4-hydroxyproline (Pro) (partial)"
 Region 128..244 /region_name="domain" /note="C-type lectin homology #label LCH"

ORIGIN 1 mslfplppll lsmvaasys etvtcedaqk tcpaviacss pgingfpgkd grdgtkgekg
 5 61 epqgqlrglq gppgklgppg npgpsgspgp kgqkgdpgks pdgdsslaas erkalqtema
 121 rikkwitfsl gkqvgnkffl tngeimtfek vkalcvkfqa svatprnaae ngaiqnlike
 181 eaflgitdek tegqfvdlgt nrlytnwne gepnnagsde dcvlllkngq wndvpctstsh
 241 lavcefp

10 SEQ ID NO: 31
 BAA86864 complement C1s [Homo sapiens] gi|6407558|dbj|BAA86864.1|[6407558]

FEATURES Location/Qualifiers source 1..329 /organism="Homo sapiens"
 /db_xref="taxon:9606" /tissue_type="peripheral leukocytes" /clone_lib="FIXII"

15 Protein 1..329 /product="complement C1s"
 CDS 1..329 /coded_by="join(AB009076.1:1142..1146,
 AB009076.1:1703..1910,AB009076.1:2118..2295,
 AB009076.1:3495..3620,AB009076.1:4328..4527,
 AB009076.1:5047..5200,AB009076.1:5748..>5863)" /note="This gene consists of
 20 total 12 exons, the last 4 exons of which were reported by Toshi, M. et al. (J. Mol. Biol.
 208:709-714, 1989)

ORIGIN 1 mwcivilfsl awvyaeptmy geilspnypq aypseveksw dievpegygi hlyfthldie
 61 lsencaydsv qiisgteeg rlcqrrssnn phspiveefq vpynklqvif ksdfsneerf
 121 tgfaayvat dinectdfvd vpcshfcnnf iggyfcscpp eyflhddmkn cgvnscgdvf
 25 181 taligeiasp nypkpypens rceyqirlek gfvvvtlrr edfdveaads agncldslvf
 241 vagdrqfpy cghgfppln ietksnaldi ifqtdltgqk kgwklyrhgd pmpcpkedtp
 301 nsvwepakak yvfrdvvqit cldgfevve

30 SEQ ID NO: 32
 CAB56124 mannose-binding lectin [Homo sapiens]
 gi|5911809|emb|CAB56124.1|[5911809]

35 FEATURES Location/Qualifiers source 1..248 /organism="Homo sapiens"
 /db_xref="taxon:9606" /chromosome="10" /map="10q11.2-q21" /note="MBL
 haplotype HYPD"

Protein 1..248 /product="mannose-binding lectin"
 sig_peptide 1..20

CDS 1..248 /gene="MBL" /coded_by="Y16582.1:892..1638"

40 ORIGIN 1 mslfplppll lsmvaasys etvtcedaqk tcpaviacss pgingfpgkd gcdgtkgekg
 61 epqgqlrglq gppgklgppg npgpsgspgp kgqkgdpgks pdgdsslaas erkalqtema
 121 rikkwitfsl gkqvgnkffl tngeimtfek vkalcvkfqa svatprnaae ngaiqnlike
 181 eaflgitdek tegqfvdlgt nrlytnwne gepnnagsde dcvlllkngq wndvpctstsh
 241 lavcefp

45 SEQ ID NO: 33
 CAB56123 mannose-binding lectin [Homo sapiens]
 gi|5911807|emb|CAB56123.1|[5911807]

50 FEATURES Location/Qualifiers source 1..248 /organism="Homo sapiens"
 /db_xref="taxon:9606" /chromosome="10" /map="10q11.2-q21" /note="MBL
 haplotype HYPA"

Protein 1..248 /product="mannose-binding lectin"
 sig_peptide 1..20

CDS 1..248 /gene="MBL" /coded_by="Y16581.1:892..1638"

ORIGIN 1 mslfpslplllsmvaasys etvtcedaak tcpaviacss pgingfpgkd grdgtkgekg
 61 epgqglrglq gppgklgppg nppsgsgppg kgqkgdpgks pdgdsslaas erkalqtema
 121 rikkwltfsl gkqvgnkffl tngeimtfek vkalcvkfqa svatprnaae ngaiqnlike
 181 eaflgitdek tegqfvdlgt nrlytnwne gepnnagsde dcvlllkngq wndvpcstsh
 241 lavcefp

SEQ ID NO: 34

CAB56122 mannose-binding lectin [Homo sapiens]

gi|5911798|emb|CAB56122.1|[5911798]

FEATURES Location/Qualifiers source 1..248 /organism="Homo sapiens"
 /db_xref="taxon:9606" /chromosome="10" /map="10q11.2-q21" /note="MBL
 haplotype LXPA"

Protein 1..248 /product="mannose-binding lectin"

sig_peptide 1..20

CDS 1..248 /gene="MBL" /coded_by="Y16580.1:892..1638"

ORIGIN 1 mslfpslplllsmvaasys etvtcedaak tcpaviacss pgingfpgkd grdgtkgekg
 61 epgqglrglq gppgklgppg nppsgsgppg kgqkgdpgks pdgdsslaas erkalqtema
 121 rikkwltfsl gkqvgnkffl tngeimtfek vkalcvkfqa svatprnaae ngaiqnlike
 181 eaflgitdek tegqfvdlgt nrlytnwne gepnnagsde dcvlllkngq wndvpcstsh
 241 lavcefp

SEQ ID NO: 35

CAB56121 mannose-binding lectin [Homo sapiens]

gi|5911796|emb|CAB56121.1|[5911796]

FEATURES Location/Qualifiers source 1..248 /organism="Homo sapiens"
 /db_xref="taxon:9606" /chromosome="10" /map="10q11.2-q21" /note="MBL
 haplotype LYPB"

Protein 1..248 /product="mannose-binding lectin"

sig_peptide 1..20

CDS 1..248 /gene="MBL" /coded_by="Y16579.1:892..1638"

ORIGIN 1 mslfpslplllsmvaasys etvtcedaak tcpaviacss pgingfpgkd grddtkgekg
 61 epgqglrglq gppgklgppg nppsgsgppg kgqkgdpgks pdgdsslaas erkalqtema
 121 rikkwltfsl gkqvgnkffl tngeimtfek vkalcvkfqa svatprnaae ngaiqnlike
 181 eaflgitdek tegqfvdlgt nrlytnwne gepnnagsde dcvlllkngq wndvpcstsh
 241 lavcefp

SEQ ID NO: 36

CAB56045 mannose-binding lectin [Homo sapiens]

gi|5911794|emb|CAB56045.1|[5911794]

/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="10"
 /map="10q11.2-q21" /note="MBL haplotype LYQC"

Protein 1..248 /product="mannose-binding lectin"

sig_peptide 1..20

CDS 1..248 /gene="MBL" /coded_by="Y16578.1:886..1632"

ORIGIN 1 mslfpslplllsmvaasys etvtcedaak tcpaviacss pgingfpgkd grdgtkeekg
 61 epgqglrglq gppgklgppg nppsgsgppg kgqkgdpgks pdgdsslaas erkalqtema
 121 rikkwltfsl gkqvgnkffl tngeimtfek vkalcvkfqa svatprnaae ngaiqnlike
 181 eaflgitdek tegqfvdlgt nrlytnwne gepnnagsde dcvlllkngq wndvpcstsh
 241 lavcefp

SEQ ID NO: 37

CAB56120 mannose-binding lectin [Homo sapiens]

gi|5911792|emb|CAB56120.1|[5911792]

5 FEATURES Location/Qualifiers source 1..248 /organism="Homo sapiens"
/db_xref="taxon:9606" /chromosome="10" /map="10q11.2-q21" /note="MBL
haplotype LYPA"

Protein 1..248 /product="mannose-binding lectin"

sig_peptide 1..20

10 CDS 1..248 /gene="MBL" /coded_by="Y16577.1:892..1638"

ORIGIN 1 msifpslpll lismvaasys etvtcedaqk tcpaviacss pgingfpgkd grdgtkgekg
61 epqgqlrglq gppgklgppg nppsgsgppg kgqkgdpgks pdgdsslaas erkalqtema
121 rikkwltfsl gkqvgnkffl tngeimtfek vkalcvkfqa svatprnaae ngaiqnlike
181 eaflgitdek tegqfvdlgt nrlytnwne gepnnagsde dcvlllknngq wndvpctstsh
15 241 lavcefpj

SEQ ID NO: 38

CAB56044 mannose-binding lectin [Homo sapiens]

gi|5911790|emb|CAB56044.1|[5911790]

20 FEATURES Location/Qualifiers source 1..248 /organism="Homo sapiens"
/db_xref="taxon:9606" /chromosome="10" /map="10q11.2-q21" /note="MBL
haplotype LYQA"

Protein 1..248 /product="mannose-binding lectin"

sig_peptide 1..20

25 CDS 1..248 /gene="MBL" /coded_by="Y16576.1:886..1632"

ORIGIN 1 msifpslpll lismvaasys etvtcedaqk tcpaviacss pgingfpgkd grdgtkgekg
61 epqgqlrglq gppgklgppg nppsgsgppg kgqkgdpgks pdgdsslaas erkalqtema
121 rikkwltfsl gkqvgnkffl tngeimtfek vkalcvkfqa svatprnaae ngaiqnlike
181 eaflgitdek tegqfvdlgt nrlytnwne gepnnagsde dcvlllknngq wndvpctstsh
30 241 lavcefpj

SEQ ID NO: 39

AAB53110 C1qR(p) [Homo sapiens] gi|2052498|gb|AAB53110.1|[2052498]

35 FEATURES Location/Qualifiers source 1..652 /organism="Homo sapiens"
/db_xref="taxon:9606" /cell_line="U937 histiocytic cell line"
Protein 1..652 /product="C1qR(p)" /function="mediates enhanced phagocytosis by
human monocytes and macrophages in response to complement C1q, mannose
binding lectin (MBL) and pulmonary surfactant protein A (SPA)"

40 CDS 1..652 /coded_by="U94333.1:149..2107" /note="C1q/MBL/SPA receptor"

ORIGIN 1 matsmgllll lllltqpga gtgadteavv cvgtacytah sgklsaaeeq nhcnqnggnl
61 atvkskeeeq hvqrvlaql rreaaltarm skfwiglqre kgkclpslp lkgfswvvgg
121 edtpysnwhk elmsciskr cvsllldlsq pllpnrpkw segpcgspgs pgsgniegfv
181 kfsfkgmcrp lalggpgqvvt yttptftss sleavpfasa anvacgegdk detqshyflc
45 241 kekapdvfdw gssgplcvsp kygcfnngg chqdcfeggd gsflcgrpg frlldlvtc
301 asmpcsssp crggatcvlg phgknytrc pqgyqldssq ldcvdvdecq dspcaqecvn
361 tpggfrcecw vgyepggpge gacqdvdeca lgrspcaqgc tndgsfhcs ceegyvlage
421 dgtqcqdvde cvgpggplcd slcfntqgsf hcglpgwvl apngvsctmg pvsigppsgp
481 pdeedkgeke gstvpraata sptrgpegtp katptsrps lssdapitsa plkmlapsgs
50 541 sgvwrepsih hataasgpqe paggdssvat qndgtdgqk lllyilgtv vaillllala
601 lgllvykrk akreekkekk pqnaadsysw vperaesram enqysptpgt dc

SEQ ID NO: 40

81

NP_571645 mannose binding-like lectin [Danio rerio]
gi|18858997|ref|NP_571645.1|[18858997]

sig_peptide 1..23

5 mat_peptide 24..251 /product="mannose binding-like lectin"
Region 24..36 /region_name="N-terminal segment"
Region 33..70 /region_name="Collagen triple helix repeat (20 copies)"
/note="Collagen" /db_xref="CDD:pfam01391"
Region 33..70 /region_name="Collagen triple helix repeat (20 copies)"
10 /note="Collagen" /db_xref="CDD:pfam01391"
Region 37..101 /region_name="collagen-like structure"
Region 37..70 /region_name="Collagen triple helix repeat (20 copies)"
/note="Collagen" /db_xref="CDD:pfam01391"
Region 71..74 /region_name="break in collagen structure"
15 Region 102..132 /region_name="neck region"
Region 133..251 /region_name="carbohydrate recognition domain" /note="CRD"
Region 134..247 /region_name="C-type lectin (CTL) or carbohydrate-recognition
domain (CRD)" /note="CLECT" /db_xref="CDD:smart00034"
Region 146..247 /region_name="Lectin C-type domain" /note="lectin_c"
20 /db_xref="CDD:pfam00059"
CDS 1..251 /gene="mbi" /coded_by="NM_131570.1:68..823" /note="collectin with
structural homology to mannose-binding lectin but with a predicted carbohydrate
specificity for galactose;mannose binding-like lectin" /db_xref="LocusID:58091"
ORIGIN 1 mallkiflga lllqlvlql magaadpql ncpayagvpg tpghnglpgr dgrvrgdgan
25 61 gpkgekgpg vnvqgppgka gppgpagakg ergpsglpgq dcmsdskse lqklsdkial
121 iekvvnfktf kkvgqkyvyt ddveetfdkg mqycssngga lvprtleen allkvfvssa
181 fkrfirtid rekegefvd drkkltftnw gnpqpdnykg aqdcgaiads glwddvscds
241 lypiiceiei k

30 SEQ ID NO: 41
BAA90338 mannose-binding lectin-associated serine protease (MASP) related
protein [Cyprinus carpio] gi|6807499|dbj|BAA90338.1|[6807499]
FEATURES Location/Qualifiers source 1..118 /organism="Cyprinus carpio"
/db_xref="taxon:7962"
35 Protein 1..118 /product="mannose-binding lectin-associated serine protease
(MASP) related protein"
CDS 1..118 /gene="MRPb"
/coded_by="join(AB030447.1:<1..96,AB030447.1:201..319,
AB030447.1:436..514,AB030447.1:616..680)" /note="MASP-related protein"
40 ORIGIN 1 kiqtgsntvs ilfhsdnsgd nlgwkltyts tgsecsplaa plnghleplq snyifkdhim
61 ltcddpgyslr qgdkefeyhq iecqrdgkws sdvplckkke sqrrhslps iltnqls

45 The second polypeptide preferably comprises at least 10, such as at least 12, for
example at least 15, such as at least 20, for example at least 25, such as at least
30, for example at least 35, such as at least 40, for example at least 50 consecutive
amino acid residues of the collectin or of a variant or a homologue to said protein.
Such a variant or homologue is preferably at least 70%, such as 80%, for example
50 90%, such as 95% identical to the collectin.

In a preferred embodiment the second polypeptide sequence comprises the CRD domain of MBL or the neck region of MBL or the collagen-like domain of MBL. More preferably the second polypeptide comprises the neck region and the CRD domain of MBL. In a most preferred embodiment the second polypeptide sequence comprises the collagen-like domain, the neck region and the CRD domain of MBL. MBL is as defined above.

Preferably the second polypeptide sequence comprises at least amino acids 170-200 of the MBL sequence shown in Figure 2, such as at least amino acids 160-200 of the MBL sequence shown in Figure 2, such as at least amino acids 150-200 of the MBL sequence shown in Figure 2, such as at least amino acids 140-200 of the MBL sequence shown in Figure 2, such as at least amino acids 130-200 of the MBL sequence shown in Figure 2, such as at least amino acids 120-200 of the MBL sequence shown in Figure 2, such as at least amino acids 110-200 of the MBL sequence shown in Figure 2, such as at least amino acids 100-200 of the MBL sequence shown in Figure 2, such as at least amino acids 90-200 of the MBL sequence shown in Figure 2, such as at least amino acids 80-200 of the MBL sequence shown in Figure 2, such as at least amino acids 70-200 of the MBL sequence shown in Figure 2, such as at least amino acids 60-200 of the MBL sequence shown in Figure 2, such as at least amino acids 80-228 of the MBL sequence shown in Figure 2.

Preferably the second polypeptide sequence comprises amino acids 80-228 of SEQ ID. NO 2.

In a preferred embodiment the second polypeptide sequence is capable of associating with at least one MASP protein, such as a MASP protein selected from the group consisting of MASP-1, MASP-2 and MASP-3 or functional homologues or variants hereof. In particular the second polypeptide is capable of associating with said at least one MASP protein when being part of the fusion protein. Thereby the second polypeptide sequence is capable of providing the fusion protein with complement system activating activity. In a preferred embodiment the second polypeptide sequence comprises an amino acid sequence selected from: 56-228 of SEQ ID. NO 2, 55-228 of SEQ ID. NO 2, 54-228 of SEQ ID. NO 2, and 50-228 of SEQ ID.

NO 2. In a preferred embodiment the second polypeptide sequence has an amino acid sequence selected from: 56-228 of SEQ ID. NO 2, 55-228 of SEQ ID. NO 2, 54-228 of SEQ ID. NO 2, and 50-228 of SEQ ID. NO 2.

- 5 In another embodiment the second polypeptide comprises the cysteine-rich region of the collectin, such as the N-terminal region of the collectin.

Fusion protein

- 10 The fusion protein comprises the first and the second polypeptide connected to each other, optionally through a linker region. In a preferred embodiment the first polypeptide sequence is positioned N-terminally in the fusion protein and the second polypeptide sequence is positioned C-terminally.

- 15 Specific examples of the components of the fusion protein are:

- A fusion protein comprising the cysteine-rich region and the collagen-like domain of L-ficolin and the CRD domain of MBL.
- 20 - A fusion protein comprising the cysteine-rich region of L-ficolin and the collagen-like domain, the neck region and the CRD domain of MBL.
- A fusion protein comprising the cysteine-rich region and the collagen-like domain of H-ficolin and the CRD domain of MBL.
- 25 - A fusion protein comprising the cysteine-rich region of H-ficolin and the collagen-like domain, the neck region and the CRD domain of MBL.
- A fusion protein comprising the cysteine-rich region and the collagen-like domain of M-ficolin and the CRD domain of MBL.
- 30 - A fusion protein comprising the cysteine-rich region of M-ficolin and the collagen-like domain, the neck region and the CRD domain of MBL.

- A fusion protein comprising the cysteine-rich region of MBL, and the CRD domain of ficolin.
- A fusion protein comprising the cysteine-rich region of MBL and the collagen-like domain, the neck region and the CRD domain of ficolin.
- A fusion protein comprising the cysteine-rich region and the collagen-like domain of L-ficolin and the CRD domain of Pulmonary surfactant-associated protein D.
- A fusion protein comprising the cysteine-rich region of L-ficolin and the collagen-like domain, the neck region and the CRD domain of Pulmonary surfactant-associated protein D.
- A fusion protein comprising the cysteine-rich region and the collagen-like domain of a ficolin and the CRD domain of a collectin-43.
- A fusion protein comprising the cysteine-rich region of a ficolin and the collagen-like domain, the neck region and the CRD domain of a collectin-43.
- A fusion protein comprising the amino acid sequence as defined by the sequence shown in Figure 3, or a functional homologue thereof, preferably a fusion protein consisting of the amino acid sequence as shown in Figure 3. In another embodiment the fusion protein has amino acid sequence 1-50 of the amino acid shown in Figure 1 and amino acid sequence 54-228 of the amino acid sequence shown in Figure 2.
- As discussed above the fusion protein is preferably capable of forming subunit complexes as well as oligomers of subunit complexes. Preferably the fusion protein forms substantially only trimeric, tetrameric, pentameric and hexameric subunit oligomers, such as trimeric, tetrameric, and pentameric subunit oligomers, such as trimeric or tetrameric subunit oligomers, more preferably substantially only tetrameric subunit oligomers, in order to obtain a more homogenous composition of fusion proteins.

Homologues

- 5 In the present context the terms homologue or variant or functional homologues are used as synonymes, wherein a homologue of a protein exhibits one or more substitutions, deletions, and/or additions of one or more amino acid residues. Fragments are a subgroup of homologues being truncations of the protein.
- 10 A homologue of the protein may comprise one or more conservative amino acid substitutions, such as at least 2 conservative amino acid substitutions, for example at least 3 conservative amino acid substitutions, such as at least 5 conservative amino acid substitutions, for example at least 10 conservative amino acid substitutions, such as at least 20 conservative amino acid substitutions, for example at least
- 15 50 conservative amino acid substitutions such as at least 75 conservative amino acid substitutions, for example at least 100 conservative amino acid substitutions. Conservative amino acid substitutions within the meaning of the present invention is substitution of one amino acid within a predetermined group of amino acids for another amino acid within the same predetermined group, exhibiting similar or substantially similar characteristics. Such predetermined groups are for example:
- 20 polar side chains (Asp, Glu, Lys, Arg, His, Asn, Gln, Ser, Thr, Tyr, and Cys,)
- non-polar side chains (Gly, Ala, Val, Leu, Ile, Phe, Trp, Pro, and Met)
- 25 aliphatic side chains (Gly, Ala Val, Leu, Ile)
- cyclic side chains (Phe, Tyr, Trp, His, Pro)
- 30 aromatic side chains (Phe, Tyr, Trp)
- acidic side chains (Asp, Glu)
- basic side chains (Lys, Arg, His)

35

amide side chains (Asn, Gln)

hydroxy side chains (Ser, Thr)

5 sulphur-containing side chains (Cys, Met), and

amino acids being monoamino-dicarboxylic acids or monoamino-monocarboxylic-monoamidocarboxylic acids (Asp, Glu, Asn, Gln).

10 Conservative substitutions may be introduced in any position of a preferred protein. It may however also be desirable to introduce non-conservative substitutions: A non-conservative substitution should lead to the formation of a homologue of a protein capable of exerting a function similar to the function of said protein. Such substitution could for example i) differ substantially in hydrophobicity, for example a hydrophobic residue (Val, Ile, Leu, Phe or Met) substituted for a hydrophilic residue such
15 as Arg, Lys, Trp or Asn, or a hydrophilic residue such as Thr, Ser, His, Gln, Asn, Lys, Asp, Glu or Trp substituted for a hydrophobic residue; and/or ii) differ substantially in its effect on polypeptide backbone orientation such as substitution of or for Pro or Gly by another residue; and/or iii) differ substantially in electric charge, for
20 example substitution of a negatively charged residue such as Glu or Asp for a positively charged residue such as Lys, His or Arg (and vice versa); and/or iv) differ substantially in steric bulk, for example substitution of a bulky residue such as His, Trp, Phe or Tyr for one having a minor side chain, e.g. Ala, Gly or Ser (and vice versa).

25 In a further embodiment the present invention relates to homologues of a preferred protein, wherein such homologues comprise substituted amino acids having hydrophilic or hydrophobic indices that are within ± 2.5 , for example within ± 2.3 , such as within ± 2.1 , for example within ± 2.0 , such as within ± 1.8 , for example within ± 1.6 , such as within ± 1.5 , for example within ± 1.4 , such as within ± 1.3 for example within ± 1.2 , such as within ± 1.1 , for example within ± 1.0 , such
30 as within ± 0.9 , for example within ± 0.8 , such as within ± 0.7 , for example within ± 0.6 , such as within ± 0.5 , for example within ± 0.4 , such as within ± 0.3 , for example within ± 0.25 , such as within ± 0.2 of the value of the amino acid it has substituted.

35

The importance of the hydrophilic and hydrophobic amino acid indices in conferring interactive biologic function on a protein is well understood in the art (Kyte & Doolittle, 1982 and Hopp, U.S. Pat. No. 4,554,101, each incorporated herein by reference).

5

The amino acid hydrophobic index values as used herein are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5) (Kyte & Doolittle, 1982).

10

The amino acid hydrophilicity values are: arginine (+3.0); lysine (+3.0); aspartate (+3.0+-.1); glutamate (+3.0+-.1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5+-.1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4) (U.S. 4,554,101).

15

Substitution of amino acids can therefore in one embodiment be made based upon their hydrophobicity and hydrophilicity values and the relative similarity of the amino acid side-chain substituents, including charge, size, and the like. Exemplary amino acid substitutions which take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

20

25

Furthermore, a homologue may comprise addition or deletion of an amino acid, for example an addition or deletion of from 2 to 100 amino acids, such as from 2 to 50 amino acids, for example from 2 to 20 amino acids, such as from 2 to 10 amino acids, for example from 2 to 5 amino acids, such as from 2 to 3 amino acids. However, additions of more than 100 amino acids, such as additions from 100 to 500 amino acids, are also comprised within the present invention.

30

Proteins sharing at least some homology with a preferred protein are to be considered as falling within the scope of the present invention when they are at least about

35

40 percent homologous, or preferably identical, with the preferred protein, such as at least about 50 percent homologous, or preferably identical, for example at least about 60 percent homologous, or preferably identical, such as at least about 70 percent homologous, or preferably identical, for example at least about 75 percent homologous, or preferably identical, such as at least about 80 percent homologous, or preferably identical, for example at least about 85 percent homologous, or preferably identical, such as at least about 90 percent homologous, or preferably identical, for example at least 92 percent homologous, or preferably identical, such as at least 94 percent homologous, or preferably identical, for example at least 95 percent homologous, or preferably identical, such as at least 96 percent homologous, or preferably identical, for example at least 97 percent homologous, or preferably identical, such as at least 98 percent homologous, or preferably identical, for example at least 99 percent homologous, or preferably identical, with the preferred protein.

Preferred proteins are complement activating proteins comprising collectins and lectins and homologues hereof.

Homologues of collectins

A homologue of a collectin including MBL within the scope of the present invention should be understood as any protein capable of exerting a function similar to the function of a collectin and comprising one or more of the variations described above. In particular such function is the ability to activate complement upon binding to one or more carbohydrates.

The terms functional homologues of collectin used herein relate to functional equivalents or a fragment of collectin comprising a predetermined amino acid sequence, and such homologues are defined as:

a) A homologue comprising an amino acid sequence capable of recognising and binding to glucans, lipophosphoglycans and glycoinositol phospholipids that contain sugar with 3- and 4-hydroxyl groups in the pyranose ring (i.e. Man, Glc, Fuc or GlcNAc) either alone or when being subunit complexed as described above and/or

- b) A homologue comprising an amino acid sequence capable of forming an association with a component of the Lectin/MBL pathway such as binding to the MASP-1, MASP-2, MASP-3 and/or sMAP either alone or when being subunit complexed as described above, wherein said binding result in activation of the Lectin/MBL pathway and/or
- c) A homologue comprising an amino acid sequence capable of by the collagen-like domain forming an oligomeric structure of two or more subunits, where a subunit comprises three identical polypeptides of a cysteine-rich region, a collagen-like domain, a neck region and a carbohydrate recognition domain.

Homologues of lectins

A homologue of a lectin including ficolins within the scope of the present invention should be understood as any protein capable of exerting a function similar to the function of a lectin and comprising one or more of the variations previously described. In particular such function is the ability to activate complement upon binding to one or more carbohydrates.

The terms functional homologues of lectin used herein relate to functional equivalents of a fragment of lectin comprising a predetermined amino acid sequence, and such homologues are defined as:

- a) A homologue comprising an amino acid sequence capable of recognising and binding to N-acetyl-glucosamine (GlcNAc), or N-acetyl-galactosamine (GalNAc), or elastin either alone or when being subunit complexed as described above and/or
- b) A homologue comprising an amino acid sequence capable of forming an association with a component of the Lectin/MBL pathway such as binding to the MASP-1, MASP-2, MASP-3 and/or sMAP either alone or when being subunit complexed as described above, wherein said binding result in activation of the Lectin/MBL pathway and/or

- c) A homologue comprising an amino acid sequence capable of by the collagen-like domain forming an oligomeric structure of two or more subunits, where a subunit comprises three identical polypeptides of a cysteine-rich region, a collagen-like domain, a neck region and a fibrinogen-like domain.

5

The activation of the lectin/MBL pathway, i.e. the activity of the fusion protein to activate the complement system may be assessed by assessing the C4 cleaving effect of the fusion protein or subunit complexes or oligomers of complexes thereof by the following method comprising the steps of

10

- applying a sample comprising a predetermined amount of fusion protein as well as a predetermined amount of MASP-1, MASP-2 or MASP-3,
- applying at least one complement factor to the sample,
- detecting the amount of cleaved complement factors,
- correlating the amount of cleaved complement factors to the amount of fusion protein, and
- determining the activity of the fusion protein.

15

20

25

The complement factor preferably used in the present method is a complement factor cleavable by the MBL/MASP-2 complex, such as C4. However, the complement factor may also be selected from C3 and C5.

30

The cleaved complement factor may be detected by a variety of means, such as by of antibodies directed to the cleaved complement factor.

The assay is carried out at conditions which minimize or eliminate interference from the classical complement activation pathway and the alternative complement activation pathway.

Preferably a homologue of a collectin and/or a lectin exhibits two of the functions defined above, more preferably three of the functions defined above.

Preparation of fusion protein

5

The fusion protein may be prepared by any suitable method known to the person skilled in the art. Below are described several of the methods for preparing the fusion protein, however the invention is not limited to those methods.

10

Synthetic preparation

15

When appropriate, in particular in relation to the size of the fusion protein, the fusion protein may be produced synthetically. The methods for synthetic production of peptides are well known in art. Detailed descriptions as well as practical advice for producing synthetic peptides may be found in Synthetic Peptides: A User's Guide (Advances in Molecular Biology), Grant G. A. ed., Oxford University Press, 2002, or in: Pharmaceutical Formulation: Development of Peptides and Proteins, Frokjaer and Hovgaard eds., Taylor and Francis, 1999.

20

Recombinant preparation

25

The fusion proteins of the invention are preferably produced by use of recombinant DNA technologies. The DNA sequence encoding each part of the fusion protein may be prepared by fragmentation of the DNA sequences encoding the full-length protein, (genomic DNA or cDNA) which the fusion protein part is derived from, using DNAase I according to a standard protocol (Sambrook et al., Molecular cloning: A Laboratory manual. 2rd ed., CSHL Press, Cold Spring Harbor, NY, 1989). The obtained DNA sequences encoding the individual parts of the fusion protein may then be fused together.

30

The DNA sequence may also be prepared by polymerase chain reaction using specific primers, for instance as described in US 4,683,202 or Saiki et al., 1988, Science 239:487-491.

The DNA sequence encoding a fusion protein of the invention may be prepared synthetically by established standard methods, e.g. the phosphoamidite method described by Beaucage and Caruthers, 1981, Tetrahedron Lett. 22:1859-1869, or the method described by Matthes et al., 1984, EMBO J. 3:801-805. According to the
5 phosphoamidite method, oligonucleotides are synthesized, e.g. in an automatic DNA synthesizer, purified, annealed, ligated and cloned in suitable vectors.

The DNA sequence is then inserted into a recombinant expression vector, which may be any vector, which may conveniently be subjected to recombinant DNA pro-
10 cedures. The choice of vector will often depend on the host cell into which it is to be introduced. Thus, the vector may be an autonomously replicating vector, i.e. a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g. a plasmid. Alternatively, the vector may be one
15 which, when introduced into a host cell, is integrated into the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

In the vector, the DNA sequence encoding a fusion protein should be operably connected to a suitable promoter sequence. The promoter may be any DNA sequence, which shows transcriptional activity in the host cell of choice and may be derived
20 from genes encoding proteins either homologous or heterologous to the host cell. Examples of suitable promoters for directing the transcription of the coding DNA sequence in mammalian cells are the SV 40 promoter (Subramani et al., 1981, Mol. Cell Biol. 1:854-864), the MT-1 (metallothionein gene) promoter (Palmiter et al., 1983, Science 222: 809-814) or the adenovirus 2 major late promoter. A suitable
25 promoter for use in insect cells is the polyhedrin promoter (Vasuvedan et al., 1992, FEBS Lett. 311:7-11). Suitable promoters for use in yeast host cells include promoters from yeast glycolytic genes (Hitzeman et al., 1980, J. Biol. Chem. 255:12073-12080; Alber and Kawasaki, 1982, J. Mol. Appl. Gen. 1: 419-434) or alcohol dehydrogenase genes (Young et al., 1982, in Genetic Engineering of Microorganisms for
30 Chemicals, Hollaender et al, eds., Plenum Press, New York), or the TPI1 (US 4,599,311) or ADH2-4c (Russell et al., 1983, Nature 304:652-654) promoters. Suitable promoters for use in filamentous fungus host cells are, for instance, the ADH3 promoter (McKnight et al., 1985, EMBO J. 4:2093-2099) or the tpiA promoter.

The coding DNA sequence may also be operably connected to a suitable terminator, such as the human growth hormone terminator (Palmiter et al., op. cit.) or (for fungal hosts) the TPI1 (Alber and Kawasaki, op. cit.) or ADH3 (McKnight et al., op. cit.) promoters. The vector may further comprise elements such as polyadenylation signals (e.g. from SV 40 or the adenovirus 5' Elb region), transcriptional enhancer sequences (e.g. the SV 40 enhancer) and translational enhancer sequences (e.g. the ones encoding adenovirus VA RNAs).

The recombinant expression vector may further comprise a DNA sequence enabling the vector to replicate in the host cell in question. An example of such a sequence (when the host cell is a mammalian cell) is the SV 40 origin of replication. The vector may also comprise a selectable marker, e.g. a gene the product of which complements a defect in the host cell, such as the gene coding for dihydrofolate reductase (DHFR) or one which confers resistance to a drug, e.g. neomycin, hygromycin or methotrexate.

The procedures used to ligate the DNA sequences coding the fusion proteins, the promoter and the terminator, respectively, and to insert them into suitable vectors containing the information necessary for replication, are well known to persons skilled in the art (cf., for instance, Sambrook et al., op.cit.).

The synthesis of the recombinant fusion protein may be by use of *in vitro* or *in vivo* cultures. The host cell culture is preferably an eucaryotic host cell culture. By transformation of an eukaryotic cell culture is in this context meant introduction of recombinant DNA into the cells. The expression construct used in the process is characterised by having the encoding region selected from mammalian genes including human genes and genes with big resemblance herewith such as the genes from the chimpanzee. The expression construct used is furthermore featured by the promoter region being selected from genes of virus or eukaryotes, including mammalian cells and cells from insects.

The process for producing recombinant MBL according to the invention is characterised in that the host cell culture is preferably eukaryotic, and for example a mammalian cell culture. A preferred host cell culture is a culture of human kidney cells and in an even more preferred form the host cell culture is a culture of human em-

bryonal kidney cells (HEK cells), such as HEK 293 cell lines for production of re-combinant human MBL. By "HEK 293 cell lines" is meant any cell line derived from human embryonal kidney tissue such as, but not limited to, the cell lines deposited at the American Type Culture Collection with the numbers CRL-1573 and CRL-10852.

Other cells may be chick embryo fibroblast, hamster ovary cells, baby hamster kidney cells, human cervical carcinoma cells, human melanoma cells, human kidney cells, human umbilical vascular endothelium cells, human brain endothelium cells, human oral cavity tumor cells, monkey kidney cells, mouse fibroblast, mouse kidney cells, mouse connective tissue cells, mouse oligodendritic cells, mouse macrophage, mouse fibroblast, mouse neuroblastoma cells, mouse pre-B cell, mouse B lymphoma cells, mouse plasmacytoma cells, mouse teratocarcinoma cells, rat astrocytoma cells, rat mammary epithelium cells, COS, CHO, BHK, VERO, HeLa, MDCK, WI38, and NIH 3T3 cells.

Alternatively, fungal cells (including yeast cells) may be used as host cells. Examples of suitable yeast cells include cells of *Saccharomyces spp.* or *Schizosaccharomyces spp.*, in particular strains of *Saccharomyces cerevisiae*. Examples of other fungal cells are cells of filamentous fungi, e.g. *Aspergillus spp.* or *Neurospora spp.*, in particular strains of *Aspergillus oryzae* or *Aspergillus niger*. The use of *Aspergillus spp.* for the expression of proteins is described in, e.g., EP 238 023.

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (for example, glycosylation) and processing (for example, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins and gene products. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. The mammalian cell types listed above are among those that could serve as suitable host cells.

Methods of transfecting mammalian cells and expressing DNA sequences introduced in the cells are described in e.g. Kaufman and Sharp, J. Mol. Biol. 159, 1982, pp. 601-621; Southern and Berg, 1982, J. Mol. Appl. Genet. 1:327-341; Loyter et al., 1982, Proc. Natl. Acad. Sci. USA 79: 422-426; Wigler et al., 1978, Cell 14:725; Corsaro and Pearson, 1981, in Somatic Cell Genetics 7, p. 603; Graham and van der Eb, 1973, Virol. 52:456; and Neumann et al., 1982, EMBO J. 1:841-845.

Other eucaryotic production systems are also envisaged by the present invention, such as the production of the fusion protein in a transgenic plant or animal.

In another aspect the present invention provides a method for producing a fusion protein by

- preparing a gene expression construct as defined above encoding a fusion protein,
- transforming a host cell culture with the construct,
- cultivating the host cell culture, thereby obtaining expression and secretion of the polypeptide into the culture medium, followed by
- obtaining a culture medium comprising recombinant fusion protein, and
- purifying the fusion protein.

The medium used to culture the cells may be any conventional medium suitable for growing mammalian cells, such as a serum-containing or serum-free medium containing appropriate supplements, or a suitable medium for growing insect, yeast or fungal cells. Suitable media are available from commercial suppliers or may be prepared according to published recipes (e.g. in catalogues of the American Type Culture Collection). Example of culture medium are RPMI-1640 or DMEM supplemented with, e.g., insulin, transferrin, selenium, and foetal bovine serum

The fusion proteins recombinantly produced by the cells may then be recovered from the culture medium by conventional procedures including separating the host

cells from the medium by centrifugation or filtration, precipitating the proteinaceous components of the supernatant or filtrate by means of a salt, e.g. ammonium sulphate, purification by a variety of chromatographic procedures, e.g. HPLC, ion exchange chromatography, affinity chromatography, or the like.

5

In a preferred embodiment the fusion protein is purified by use of carbohydrate affinity chromatography as described above. In a preferred embodiment of the invention the affinity chromatography is performed by means of matrices of mannose, hexose or N-acetyl-glucosamine derivatized matrices, which are suitable for affinity chromatography. In particular, an affinity chromatography is used, in which the matrices have been derivatized with mannose.

10

Purified recombinant fusion protein is in this context to be understood as recombinant fusion protein purified from cell culture supernatants or body fluids or tissue from transgenic animals purified by use of for example carbohydrate affinity chromatography.

15

After application of the culture media the column is washed, preferably by using non-denaturing buffers, having a composition, pH and ionic strength resulting in elimination of proteins, without eluting the fusion protein. Such a buffer may be TBS. Elution of fusion protein is performed with a selective desorbing agent, capable of efficient elution of fusion protein, such as TBS containing a desorbing agent, such as EDTA (5 mM for example) or mannose (50 mM for example), and fusion proteins are collected.

20

25

Pharmaceutical composition and treatment

30

The fusion protein obtained by the present invention may be used for the preparation of a pharmaceutical composition for the prevention and/or treatment of various diseases or conditions. In the present context the term pharmaceutical composition is used synonymously with the wording medicament.

35

In addition to the fusion protein, the pharmaceutical composition may comprise a pharmaceutically acceptable carrier substance and/or vehicles.

In particular, a stabilising agent may be added to stabilise the fusion proteins. The stabilising agent may be a sugar alcohol, saccharide, protein and/or amino-acids. An example of a stabilising agent may be albumin or maltose.

- 5 Other conventional additives may be added to the pharmaceutical composition depending on administration form for example.

In one embodiment the pharmaceutical composition is in a form suitable for injections. Conventional carrier substances, such as isotonic saline, may be used.

10

In another embodiment the pharmaceutical composition is in a form suitable for pulmonal administration, such as in the form of a powder for inhalation or creme or fluid for topical application.

15

A treatment in this context may comprise cure and/or prophylaxis of e.g. the immune system and reproductive system by humans and by animals having said functional units acting in this respect like those in humans. By conditions to be treated are not necessarily meant conditions presently known to be in a need of treatment, but comprise generally any condition in connection with current and/or expected need or

20

in connection with an improvement of a normal condition. In particular, the treatment is a treatment of a condition of deficiency of lectins, such as MBL deficiency.

In another aspect of the present invention the manufacture is provided of a medication consisting of said pharmaceutical compositions of fusion protein intended for

25

treatment of conditions comprising cure and/or prophylaxis of conditions of diseases and disorders of e.g. the immune system and reproductive system by humans and by animals having said functional units acting like those in humans.

Said diseases, disorders and/or conditions in need of treatment with the compounds of the invention comprise eg treatment of conditions of deficiency of MBL, treatment of cancer and of infections in connection with immunosuppressive chemotherapy including in particular those infections which are seen in connection with conditions during cancer treatment or in connection with implantation and/or transplantation of organs. The invention also comprises treatment of conditions in connection with

30

35

recurrent miscarriage.

Thus, in particular the pharmaceutical composition may be used for the treatment and/or prevention of clinical conditions selected from infections, MBL deficiency, cancer, disorders associated with chemotherapy, such as infections, diseases associated with human immunodeficiency virus (HIV), diseases related with congenital or acquired immunodeficiency. More particularly, chronic inflammatory demyelinating polyneuropathy (CIDP), Multifocal motoric neuropathy, Multiple sclerosis, Myasthenia Gravis, Eaton-Lambert's syndrome, Opticus Neuritis, Epilepsy; Primary antiphospholipid syndrome; Rheumatoid arthritis, Systemic Lupus erythematosus, Systemic scleroderma, Vasculitis, Wegner's granulomatosis, Sjögren's syndrome, Juvenile rheumatoid arthritis; Autoimmune neutropenia, Autoimmune haemolytic anaemia, Neutropenia; Crohn's disease, Colitis ulcerous, Coeliac disease; Asthma, Septic shock syndrome, Chronic fatigue syndrome, Psoriasis, Toxic shock syndrome, Diabetes, Sinuitis, Dilated cardiomyopathy, Endocarditis, Atherosclerosis, Primary hypo/agammaglobulinaemia including common variable immunodeficiency, Wiskot-Aldrich syndrome and serve combined immunodeficiency (SCID), Secondary hypo/agammaglobulinaemia in patients with chronic lymphatic leukaemia (CLL) and multiple myeloma, Acute and chronic idiopathic thrombocytopenic purpura (ITP), Allogenic bone marrow transplantation (BTM), Kawasaki's disease, and Guillan-Barre's syndrome.

The route of administration may be any suitable route, such as intravenously, intramuscularly, subcutaneously or intradermally. Also, pulmonal or topical administration is envisaged by the present invention.

In particular the fusion protein may be administered to prevent and/or treat infections in patients having clinical symptoms associated with congenital or acquired MBL deficiency or being at risk of developing such symptoms. A wide variety of conditions may lead to increased susceptibility to infections in MBL-deficient individuals; such as chemotherapy or other therapeutic cell toxic treatments, cancer, AIDS, genetic disposition, chronic infections, and neutropenia.

The pharmaceutical composition may thus be administered for a period before the onset of administration of chemotherapy or the like and during at least a part of the chemotherapy.

The fusion protein may be administered as a general "booster" before chemotherapy, or it may be administered to those only being at risk of developing MBL deficiency. The group of patients being at risk may be determined by measuring the MBL level before treatment and only subjecting those to treatment whose MBL level is below a predetermined level.

The fusion protein is administered in suitable dosage regimes, in particular it is usually administered at suitable intervals, eg. once or twice a week during chemotherapy.

Normally from 1-100 mg is administered per dosage, such as from 2-10 mg, mostly from 5-10 mg per dosage. Mostly about 0.1 mg/kg body weight is administered.

Furthermore, an aspect of the present invention is the use of a recombinant composition according to the present invention in a kit-of-parts further comprising another medicament. In particular the other medicament may be an anti-microbial medication, such as antibiotics.

Concerning miscarriage, it has been reported that the frequency of low plasma levels of MBL is increased in patients with otherwise not explained recurrent miscarriages, which is the background for lowering of the susceptibility to miscarriage by a reconstitution of the MBL level by administration of recombinant MBL in these cases.

As to the nature of compounds of the invention, it appears, that in its broad aspect, the present invention relates to compounds which are able to act as opsonins, that is, able to enhance uptake by macrophages either through direct interaction between the compound and the macrophage or through mediating complement deposition on the target surface.

Examples

Example 1

Plasmidcloning of FCNMBL-r1, -r2, -r3,-r4,-r5,-r6 and-r7.**1.1 Summary**

A series of plasmids were constructed for the expression in mammalian cells of protein fusions between recombinant human mannose-binding lectin 2 gene (rhMBL) and human ficolin 2 (FCN2). The vector is derived from a high-copy-number ColE1-based plasmid and is designed to allow protein expression in mammalian systems. The fusion protein expressions are driven by the human cytomegalovirus (CMV) immediate early promoter to promote constitutive expression. Selection is made possible in bacteria by the ampicillin-resistance gene under control of the prokaryotic β -lactamase promoter. The neomycin-resistance gene is driven by the SV40 early promoter, which provides stable selection with G418 in mammalian cells.

1.2 Constructs and experimental work

In order to express fusion proteins between Ficolin2 and MBL we have designed and constructed a series of plasmids. The new recombinant plasmids are based on the previously cloned pcDNA2001-cintMBLcDNA. This plasmid contains a synthetic intron together with the cDNA for human MBL.

The following fusions were designed (underlined font indicates FCN2 part - *italics* indicate MBL part of the fusion protein.)

FCN2MBLr1 (SEQ ID NO:118):

FCN2 (signalseq+ collagen+"hinge" to ficolin dom aa131) MBL (from aa129 carbohydrate bind dom.)

MELDRAVGVLGAATLLLSFLGMAWALQAADTCPEVKMVGLEGSDKLTILRGCP-GLPGAPGDKGEAGTNGKRGERG

PPGPPGKAGPPGPNGAPGEPQPCLTGPRTCKDLLDRGHFLSGWHTIYLPDCR-PLTFSLGKQVGNGKFFLTNGEIMT

FEKVKALCVKFQASVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLTGN-RLTYTNWNEGEPNNAGSDEDC

VLLLKNGQWNDVPCSTSHLAVCEFPI

FCN2MBLr2 (SEQ ID NO: 119):

101

FCN2 (signalseq+ collagen+"hinge"+part of ficolin-dom. containing pred. coil-coil to aa207) MBL (from aa129 carbohyd.bind dom.)

5 MELDRAVGVLGAATLLLSFLGMAWALQAADTCPEVKMVGLEGSDKLTILRGCP-
GLPGAPGDKGEAGTNGKRGGERG

PPGPPGKAGPPGPNAPGEPQPCLTGPRTCKDLLDRGHFLSGWHTIYLPDCR-
PLTVLCDMDTDGGGWTVFQRRVD

10 GSVDFYRDWATYKQGFGSRLGEFWLGNDNIHALTAQGTSELRVDLVDFEDNY-
QFAKLTFSLGKQVGNKFFLTNGE

*IMTFEKVKALCVKFQASVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLT-
GNRLTYTNWNEGEPNNAGSD*

15 *EDCVLLLKNGQWNDVPCSTSHLAVCEFP*

FCN2MBLr3 (SEQ ID NO: 120):

20 FCN2 (signalseq+ collagen to aa92) MBL (from aa101 coil-coil + carbohyd.bind dom.)

25 MELDRAVGVLGAATLLLSFLGMAWALQAADTCPEVKMVGLEGSDKLTILRGCP-
GLPGAPGDKGEAGTNGKRGGERG

PPGPPGKAGPPGPNAPDGDSSLAASERKALQTEMARIKKWLTFSLG-
KQVGNKFFLTNGEIMTFEKVKALCVKF

30 *QASVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLTGNRLTYTN-
WNEGEPNNAGSDEDCVLLLKNGQWND*

VPCSTSHLAVCEFP

35 **FCN2MBLr4 (SEQ ID NO: 121):**

40 FCN2 (signalseq+ part of collagen to cons.K at aa93) MBL (from cons.K at aa77 rest of collagen+coil-coil + carbohyd.bind dom.)

MELDRAVGVLGAATLLLSFLGMAWALQAADTCPEVKMVGLEGSDKLTILRGCP-
GLPGAPGDKGEAGTNGKRGGERG

45 PPGPPGKLGPPGNPGPSGSPGPKGQKGDGKSPDGDSSLAASERKALQTEMA-
RIKKWLTFSLGKQVGNKFFLTNG

*EIMTFEKVKALCVKFQASVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLT-
GNRLTYTNWNEGEPNNAGS*

50 *DEDCVLLLKNGQWNDVPCSTSHLAVCEFP*

FCN2MBLr5 (SEQ ID NO: 122):

FCN2 (signalseq+ part of collagen to cons.G at aa69) MBL (from cons.G at aa.64 rest of collagen(containing "kick")+coil-coil + carbohyd.bind dom.)

5 MELDRAVGVLGAAATLLLSFLGMAWALQAADTCPEVKMVGLEGSDKLTILRGCP-
GLPGAPGDKGEAGTNGQGLRGL

QGPPGKLGPPGNPGPSGSPGPKGQKGDPGKSPDGDSSLAASERKALQTEMA-
RIKKWLTFSLGKQVGNKFFLTNGE

10 IMTFEKVKALCVKFQASVATPRNAAENGAIQNLIKEEAF LGITDEKTEGQFVDLT-
GNRLTYTNWNEGEPN NAGSD

EDCVLLLKNGQWNDVPCSTSHLAVCEFPI

15

FCN2MBLr6 (SEQ ID NO: 123):

20 MBL (replaced MBLcollagen(aa.41 to aa 99)+coil-coil + carbohyd.bind dom.) FCN2
(inserted collagen aa.54 to aa.92)

MSLFPSLP LLLLSMVAASYSETVTCEDAQKTCPAVIACSSPGCPGLPGAPGDK-
GEAGTNGKRGERGPPGPPGKAG

25 PPGPNGAPSPDGDSSLAASERKALQTEMARIKKWLTFSLGKQVGNKFFLT-
NGEIMTFEKVKALCVKFQASVATPR

NAAENGAIQNLIKEEAF LGITDEKTEGQFVDLTGNRLTYTNWNEGEPN NAGSDED-
CVLLLKNGQWNDVPCSTSHL

30

AVCEFPI

FCN2MBLr7 (SEQ ID NO: 124):

35

MBL (signal seq. to aa.25)FCN2 (collagen to aa93) MBL (from aa100 coil-coil + car-
bohyd.bind dom.)

40 MSLFPSLP LLLLSMVAASYSALQAADTCPEVKMVGLEGSDKLTILRGCPGLPGAP-
GDKGEAGTNGKRGERGPPGP

PGKAGPPGPNGAPSPDGDSSLAASERKALQTEMARIKKWLTFSLGKQVGNKF-
FLTNGEIMTFEKVKALCVKFQAS

45 VATPRNAAENGAIQNLIKEEAF LGITDEKTEGQFVDLTGNRLTYTNWNEGEPN-
NAGSDEDCVLLLKNGQWNDVPC

STSHLAVCEFPI

Parental plasmids used for all constructions :

- pcDNA2003-cintMBLcDNA
- Invitrogen Genestorm clone RG000632 (Cat. No. H-K1000 Invitrogen).

5 Constructions were done by recombination using the BD In-Fusion™ PCR Cloning Kit from BD (Cat. No. 631774). The BD In-Fusion Kit allows the cloning of PCR products based only on 2 x 15 bp homology between vector and end of the PCR product. Ligase, or phosphatase are unnecessary when cloning with this kit. The In-Fusion enzyme captures the DNA fragment ends and fuses the insert to the vector.

10 Primers used for the PCR reactions are shown in table 1.

PCR reactions and linearization of vector for recombination

PCR reactions were done on plasmid "Genestorm RG000632" batch N135-15C digested with Bstz17I (N135-20B). Primers pairs were used as described below. Kit for PCR reactions : PfuUltra™ Hotstart PCR Master Mix Stratagene #600630. The PCR reaction tubes were run on the BioRAD i-cycler using the temperature profile shown in table 2.

For the recombination reactions the vector pcDNA2001-cintMBLcDNA was linearized by restriction enzyme digestion with the enzymes listed below.

20

FCN2MBLr1:

PCR using primers : Pr1-xho-MBLFCN + Pr4-Xmn-FCNMBL-rev (product 463 bp)
Digest of pcDNA2001-cintMBLcDNA : XhoI + XmnI (partial)

25

FCN2MBLr2:

PCR USING PRIMERS : Pr1-xho-MBLFCN + Pr5-Xmn-FCNMBL-rev
Digest of pcDNA2001-cintMBLcDNA : XhoI + XmnI (partial)

30

FCN2MBLr3:

PCR USING PRIMERS : Pr1-xho-MBLFCN + Pr6-b-Bsp-FCNMBL-rev
Digest of pcDNA2001-cintMBLcDNA : XhoI + BspEI

FCN2MBLr4:

PCR USING PRIMERS : Pr1-xho-MBLFCN + Pr2-apa-FCNMBL-rev

Digest of pcDNA2001-cintMBLcDNA : XhoI + ApaI

FCN2MBLr5:

5 PCR USING PRIMERS : Pr1-xho-MBLFCN + Pr3-apa-FCNMBL-rev

Digest of pcDNA2001-cintMBLcDNA : XhoI + ApaI

FCN2MBLr6:

PCR USING PRIMERS : Pr8-BstAP-MBLFCN + Pr6-b-Bsp-FCNMBL-rev

10 Digest of pcDNA2001-cintMBLcDNA : partial BstAPI + BspEI

FCN2MBLr7:

PCR USING PRIMERS : Pr7-Alw-MBLFCN + Pr6-b-Bsp-FCNMBL-rev

15 Digest of pcDNA2001-cintMBLcDNA : partial AlwNI + BspEI

In-Fusion PCR recombination reactions

In-Fusion PCR recombination reactions were set up using approx. 50–100 ng of Quiagen Minelute purified PCR products together with 50–100 ng of Quiagen Minelute purified linearized pcDNA2001-cintMBLcDNA .

20 1/10 of the recombination reactions were transformed into MAX efficiency DH5α Competent Cells (Invitrogen Cat. No. 18258-012). 1/10 and 9/10 from each transformation were spread on separate LB plates containing 200 ug/ml ampicillin. Plates were incubated at 37°C overnight.

25 Screening for positive clones : At least 6 colonies from each experimental plate were picked for miniprep plasmid DNA isolation. To determine the presence of insert, DNA was analyzed by restriction digest analysis with the enzyme *Pst*I. Three individual positive clones from each reaction were chosen for further work.

Restriction Analysis

30 In order to verify the selected individual recombinant plasmids after the primary screen we performed an intensive restriction enzyme digestion analysis on the plasmid DNA isolated.

Plasmid DNA of the recombinants were digested with the enzyme shown in table 3. The expected fragments are also listed in the table. All recombinant clones tested exhibited the expected pattern. Digestion with *Eco*RI was not as predicted. An addi-

tional fragment was observed both in digestion of the recombinant as well of the parental plasmid. The discrepancy can be explained by an additional EcoR1 site on the parental plasmid.

Results

- 5 Recombinant plasmids obtained are shown schematically in figures 4-8 for constructs r1, -r2, -r3, -r4, and -r5.

Example 2

- 10 **Experiments with transient expression of recombinant fusion proteins of human MBL and human FCN2**

2.1 Summary

- 15 We report the expression of recombinant human fusion proteins FCNMBLr1, FCNMBLr4, FCNMBLr5 and MBL in HEK293 and Per.C6 cells. We found that the cell lines in the transient transfection experiment were able to produce at least the fusion proteins FCNMBLr4 and FCNMBLr5 assembled in active oligomers with a structure primarily similar to MBL oligomer forms 3 and 4. The fusion proteins FCNMBLr4 and FCNMBLr5 behaved like MBL upon binding to a carbohydrate surface and upon activating the complement cascade.
- 20

2.2 Introduction

- 25 The aim of the studies was to elucidate the possibility of creating a hybrid protein consisting of the collagen part of human ficolin 2 and the human mannose binding lectin (MBL). Furthermore we wished to clarify if such molecules would still possess the ability to bind to complex carbohydrate structures and still are able to activate complement.

- Two eukaryotic cell lines of human origin HEK293 and Per.C6 were used as host cell lines for transient transfections with the respective expression plasmids. Transcription was driven by the CMV-IE promoter enhancer.
- 30

2.3 Experimental

Material and Methods

Plasmids used for the transfection experiments

pME607-FCNMBL-r1, -r2, -r3, -r4, -r5, -r6 and -r7 (described in example 1)

Origin of Cells used

PerC6 cells were obtained from Crucell.

HEK 293 Freestyle cells were obtained from InVitrogen.

Culture media

- 5 PerC6 cells were cultured at 37°C in 10% (vol/vol) CO₂ maintained as monolayers in serum free medium.

HEK 293 Freestyle were cultured at 37°C in 8% (vol/vol) CO₂ maintained as suspension in an InVitrogen Freestyle medium.

Transfections and harvest of media

- 10 Per.C6 cells in serum containing medium were transfected with the DNA using the transfection reagent Lipofectamine. One day after transfection the medium was replaced with serum free medium.

- HEK293 cells in serum free Freestyle medium were transfected with the DNA using the transfection reagent 293fect. The medium was collected after approximately 4
15 days of incubation after transfection.

Quantification of MBL

Recombinant MBL assay (TRIFMA) using Mannan coated plates or mAb-131-01 coated plates. For quantification of MBL, time-resolved immunofluorometry was carried out.

- 20 SDS-PAGE and Western blot analysis

SDS-PAGE with subsequent electrophoretic transfer of proteins to polyvinylidene difluoride membranes and detection of MBL using monoclonal anti-MBL antibody was carried out.

C4 assay

- 25 The assay is designed to measure MBL and rMBL abilities to initiate the MBL Lectin-pathway of the complement system. MBL associated serine protease (MASP 2) associated with MBL cleaves the complement factor C4 releasing C4a and C4b. The C4b deposition on the Mannian coated ELISA plates is detected with biotin labelled antibodies against C4b and Europium labelled Streptavidin.

- 30 **2.4 RESULTS**

In the experiments described herein we were able to express FCN2MBLr4, FCN2MBLr5 and MBL transiently in both HEK293 and Per.C6 cells under serum free conditions.

Oligomeric form of the fusion proteins

The oligomeric forms of the fusion protein were examined by non-reducing denaturing SDS PAGE followed by a western blot. The detecting antibody recognizes the CDR part of MBL (and maybe part of the coil-coil region). The results are shown in figure 10. It is evident from the figure that the most prominent form of the fusion proteins FCN2MBLr4 and FCN2MBLr5 is approximately 250 kDa corresponding to a 3- or 4-mere of subunits consisting of 3 single protein chains (24 kDa). The appearance of the oligomeric form was independent of the host cells used. MBL was produced in a wide range of oligomeric forms.

Binding properties

The fusion proteins were tested for functionality of the MBL carbohydrate binding domain by binding to a mannan surface and detection with an antibody that recognizes the CDR part of MBL (and maybe part of the coil-coil region). The results are shown in table 4. It can be concluded that FCN2MBLr4 and FCN2MBLr5 were expressed just as well as MBL in the host cells and that the fusion proteins bind to a mannan surface.

MASP-2 binding and C4 cleavage

The fusion proteins were further tested for the capacity to bind MASP-2 and for activating the serine protease of MASP-2. This was done by measuring cleavage of the MASP-2 substrate complement factor C4 upon binding of the fusion protein to a mannan surface. Results are shown in table 5. It can be concluded from these results that the fusion proteins FCN2MBLr4 and FCN2MBLr5 preserved the ability to bind and activate MASP-2.

Discussion

The results described herein clearly demonstrate that it is possible to construct fusion proteins of FCN 2 and MBL with the following properties:

1. The oligomeric structure of the fusion proteins is more simple than that of the MBL protein.
2. The fusion proteins keep the essential property of MBL activation of the complement cascade upon binding to a dense carbohydrate structure.

Table 1. Primers used for the PCR reactions

Sequence typed in bold shows the 15 bp homology needed for the recombination into the vector.

Primer name	DNA Sequence of Primer	Primer part of pcDNA2001-cintMBLcdna	Primer part of RG000632
Pr1-xho-MBLFCN	ataggctagcctcgaagctcgccctcaccatg-gagctggacag	ataggctagcctcga	agctcgccctcaccatggagctggacag
Pr2-apa-FCNMBL-rev	Ccaactttccaggggggcccggggggccacgttctcctctcttcc	g(replaced) ggccccctggaaagtgg	ggaaagagaggagaacgtggcccccc
Pr3-apa-FCNMBL-rev	Ccaactttccaggggggcccgtgaagcctctgagccctgtccattggtgctgctctccctggg	Caagggtca-gaggctta-cagggccccctgga aagtgg	cccaaggga-gaggcaggcac-caatgga
Pr4-Xmn-FCNMBL-rev	Tggtcaggaagaactgttcccaactgtttgcc-cagagagaaagt-caggggccggcagtcgggcagg	Ttctctctgggcaaa-caagtgggaa- caagttcttctgacc a	cctgccgactgccggccctgact
Pr5-Xmn-FCNMBL-rev	Tggtcaggaagaactgttcccaactgtttgcc-cagagagaactag-caaactggtagtgtcctcaaagtcc	Ttctctctgggcaaa-caagtgggaa- caagttcttctgacc a	ggacttgaggacaactac-cagttgctaag
Pr6-b-Bsp-FCNMBL-rev	Gactatcaccatccggaggtgctcgtggggcc-caggtgtcc	ccggatggtgatagt	ggac-cacctggggccaacggagcacct
Pr7-Alw-MBLFCN	Cagcgtcttactcagctctccaggcggcagacacctgtcc	cagcgtcttactcag	ctctccaggcggcagacacctgtcc
Pr8-BstAP-MBLFCN	Agacctgccctgcagtgattgcctgtagctctc-caggctgtccggggctgctggggcccc	Agacctgccctgca gtgattgcctgtagctct cca	ggctgtccggggctgctggggcccc

Table 2:

Cycle	times	step	Temp	Time
1	1x	1	95°	21 min
2	30x	1	95°	0 min 30 sec
		2	72° (67.5° for r2)	0 min 30 sec
		3	72°	1 min
3	1x	1	72°	10 min
4	1x	1	4°	∞

5

Table 3:

pcDNA2001-cintMBLcdNA	pME607-FCN2M BLr1	pME607-FCN2M BL r2	pME607-FCN2M BL r3	pME607-FCN2M BL r4	pME607-FCN2M BL r5
<i>Pst</i> 4212	<i>Pst</i> 4212	<i>Pst</i> 4212	<i>Pst</i> 4212	<i>Pst</i> 4212	<i>Pst</i> 4212

1586 405 375	1586 788	1586 1016	1486 782	1586 800	1586 797
<u>EcoRI</u> 5810 768	<u>EcoRI</u> 6586	<u>EcoRI</u> 6814	<u>EcoRI</u> 6580	<u>EcoRI</u> 6598	<u>EcoRI</u> 6595
<u>XmaI</u> 6578	<u>XmaI</u> 6586	<u>XmaI</u> 6814	<u>XmaI</u> 6580	<u>XmaI</u> 6589	<u>XmaI</u> 6595
<u>BstXI</u> undigested	<u>BstXI</u> 6586	<u>BstXI</u> 6814	<u>BstXI</u> 6580	<u>BstXI</u> 6598	<u>BstXI</u> 6595
<u>BstAPI</u> 4622 1469 415 72	<u>BstAPI</u> 4622 1892 72	<u>BstAPI</u> 4622 2120 72	<u>BstAPI</u> 4622 1886 72	<u>BstAPI</u> 4622 1904 72	<u>BstAPI</u> 4622 1901 72
<u>NcoI</u> 3435 2408 735	<u>NcoI</u> 3435 1747 735 669	<u>NcoI</u> 3435 1975 735 669	<u>NcoI</u> 3435 1741 735 669	<u>NcoI</u> 3435 1759 735 669	<u>NcoI</u> 3435 1756 735 669

Table 4 MBL binding to mannan measured by TRIFMA

		$\mu\text{g MBL equivalents /ml}$
FCN2MBLr5	HEK293	0,689
FCN2MBLr5	HEK293	0,764
FCN2MBLr1	HEK293	0,874
MBL	HEK293	0,457
FCN2MBLr4	HEK293	0,851
MBL	HEK293	1,885
FCN2MBLr5	Per.C6	0,296
MBL	Per.C6	0,271
FCN2MBLr4	Per.C6	0,077
FCN2MBLr4	Per.C6	0,091
FCN2MBLr4	Per.C6	0,089
FCN2MBLr4	Per.C6	0,035
MBL	Per.C6	0,092

5 Table 5 C4 activity of the fusion proteins

	Cells	Aktivitet +/-
pME607-FCNMBLr5 clone 1	HEK293	+
pME607-FCNMBLr5 clone 5	HEK293	+
pME607-FCNMBLr4 clone 2	HEK293	+/+ (after purification)
pcDNA2001-cintMBLcDNA	HEK293	+
pME607-FCNMBLr5 clone 5	Per.C6	+
pME607-FCNMBLr4 clone 2 (maxi)	Per.C6	-/(+) (after purification)